



Pathos in Natural Language Argumentation: Emotional Appeals and Reactions

Barbara Konat¹ · Ewelina Gajewska¹ · Wiktoria Rossa¹

Accepted: 22 February 2024 / Published online: 21 June 2024
© The Author(s) 2024

Abstract

In this paper, we present a model of pathos, delineate its operationalisation, and demonstrate its utility through an analysis of natural language argumentation. We understand pathos as an interactional persuasive process in which speakers are performing pathos appeals and the audience experiences emotional reactions. We analyse two strategies of such appeals in pre-election debates: pathotic Argument Schemes based on the taxonomy proposed by Walton et al. (Argumentation schemes, Cambridge University Press, Cambridge, 2008), and emotion-eliciting language based on psychological lexicons of emotive words (Wierzba in Behav Res Methods 54:2146–2161, 2021). In order to match the appeals with possible reactions, we collect real-time social media reactions to the debates and apply sentiment analysis (Alswaidan and Menai in Knowl Inf Syst 62:2937–2987, 2020) method to observe emotion expressed in language. The results point to the importance of pathos analysis in modern discourse: speakers in political debates refer to emotions in most of their arguments, and the audience in social media reacts to those appeals using emotion-expressing language. Our results show that pathos is a common strategy in natural language argumentation which can be analysed with the support of computational methods.

Keywords Natural language argumentation · Pathos appeals · Emotional reactions · Emotion-eliciting language · Pathotic argument schemes · Expressing emotions in language

✉ Barbara Konat
bkonat@amu.edu.pl

Ewelina Gajewska
ewegaj@st.amu.edu.pl

Wiktoria Rossa
wikros1@st.amu.edu.pl

¹ Faculty of Psychology and Cognitive Science, Adam Mickiewicz University in Poznan, Szamarzewskiego St. 89 AB, 60-568 Poznań, Poland

1 Introduction

Appeals to emotions have accompanied argumentation since the dawn of rhetoric. Following the introduction of the Aristotelian triad of *logos*, *ethos*, and *pathos* (Aristotle 2004), it has been recognised that the persuasive power of public speech can derive not only from the strength of the arguments or the trustworthiness of the rhetor but also from the ability to elicit, ignite, and regulate the emotions of the audience.

In modern argumentation studies, the place of emotion has been theoretically established through seminal works of Walton (1992), Gilbert (2004) and others (Micheli 2010; Plantin 2019). These scholars have laid the conceptual groundwork and have primarily focused on theoretical considerations or detailed linguistic analyses. Recent advances allow to capture pathos and emotions both in natural language annotation (Hidey et al. 2017) and in psychological experiments (Villata et al. 2017). Contributing to this landscape, this paper introduces a novel model and methodology. Our approach advances previous work in two significant ways: first, at the theoretical level, we propose a clear distinction between the appeal to emotions on the part of the speaker and the emotional reactions experienced by the audience. Second, we offer a methodology and tools from both psychology and computational linguistics, accompanied by an empirical evaluation of these tools within the context of natural language argumentation. As a material for pathos analysis, we use pre-election debates, such as the following example from a debate preceding Polish parliamentary election, by a representative of the left-wing party Razem, Adrian Zandberg:

Example 1 Adrian Zandberg (Oct 10, 2019; TVP):

Conclusion: *We should quit coal by 2035. It is not the question of ambition, but of elementary **responsibility**.*

Premise: *Because the climate **crisis** is not, Mr, Bosak, an ideology. It is what practically all **scientists** have been telling us: human impact on the climate will mean, for our **children**'s and **grandchildren**'s generations, **dramatic** problems. It will not only mean drought in summer, not only **higher** food costs, not only **the lack** of water in the cities, not only heat waves, and the fact that dozens of thousands of elderly will **die**. It is a **danger** for the civilisation itself.*

Intuitively, the reader may perceive the emotional load of this argument as well as the balance it introduces between positive representations (words such as “children” and “grandchildren”), trust building (“scientists”), and fear stirring (“dramatic”, “danger”, “death”). In this study, we aim to develop a theoretical understanding of the manner in which strategies such as the one illustrated in Example 1 are used, and the emotions they elicit, by carefully analysing the *logos* and *pathos* of pre-election debates. In *Rhetoric* (Book I, 2), Aristotle (2004) writes: “persuasion may come through the listeners, when the speech stirs their emotions. Our judgements when we are pleased and friendly are not the same as when we are pained and hostile”. Following this approach, we present an analysis of pathos in natural language

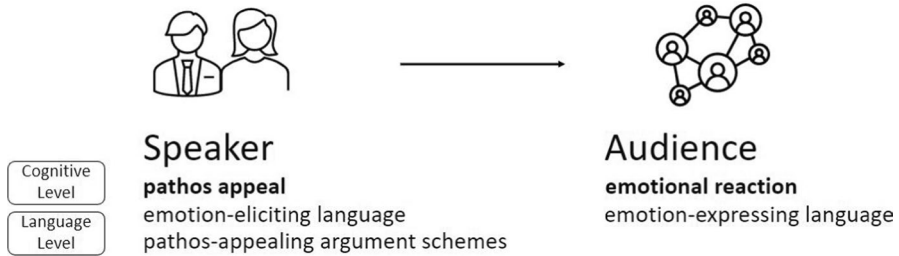


Fig. 1 Model of pathos and its operationalisation

argumentation by proposing a model and method for capturing both emotional appeals and reactions. We observe how speakers appeal to emotions in two ways: first, by using pathotic Argument Schemes [for example Fear Appeal, as described by Walton (2013)], second, by using emotion-eliciting language [for example words such as “war” or “children”, see Wierzba et al. (2021)]. On the side of the audience we capture the emotional reactions expressed in language using sentiment analysis methods [see Alswaidan and Menai (2020)].

We understand pathos as an interactional event, of pragmatic and persuasive nature. Pathos occurs when the speaker is intentionally attempting at eliciting emotions in the audience for the persuasive gain, using linguistic means. These strategies can cause emotional reaction in the listener, who, in turn, can use emotion-expressing language. This entails that pathos can occur only adjacent to logos, i.e. we analyse only those appeals to emotions which appear within argumentation.

In order to capture both speaker’s appeals and audience’s reactions, we propose a new model of pathos, suitable for modern rhetoric and new discourse genres of the digital era. The model, presented in Fig. 1 integrates cognitive and linguistic dimensions, by observing how intention of eliciting emotions (pathos appeals) on the side of the speaker results in the emotional reactions of the audience. Those cognitive processes can be observed on a language level using automated and semi-automated methods.

The model proposed here is of course highly idealised, as it neglects several important factors of real-life persuasion. The speaker can influence an audience in more ways, using non-verbal communication or ethos appeals. The audience in turn can (and probably will) have some pre-existing emotional states influencing their perception. Finally, appeals to emotions will occur not only within argumentation but also in other parts of dialogue, and—respectively—the audience will express their emotion not only as a reaction to speaker’s appeal. All those factors are neglected in the proposed model. We believe, however, that there is a strong gain from such idealisation - the proposed model allows for operationalisation of pathos on computational level, paving the way for systematic empirical analysis of emotions in argumentation.

To address the subject of emotional appeals in natural language argumentation, we propose a unique combination of three methodologies presented in Fig. 2. First, from argumentation theory, we use an established method of manual annotation, that

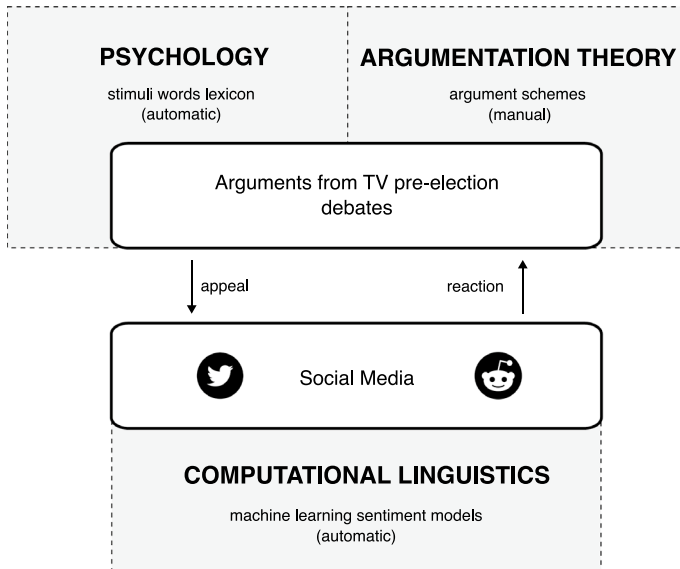


Fig. 2 Summary of methodology and data used in the study

is employing human analysts for coding the cases found in natural language. For this study, we select eight Argument Schemes that we consider to be pathos-related, following the framework presented by Walton et al. (2008), and mark their appearances in the transcripts of pre-election debates from Poland and the United States. For the second method, to capture the use of emotion-eliciting language in arguments, we adopt the approach from experimental psychology (Mohammad and Turney 2010; Wierzba et al. 2015, 2021), in which a given word can be seen as a stimuli (similarly to picture or a sound) igniting selected emotion in response. We use the lexicons created by psychologists, in order to automatically mark the presence of emotion-eliciting words in premises and conclusions of natural language arguments. These two methods allow us to capture the two ways of appealing to emotions with the means of language. The third method relates to the audience response. How effective are pathos appeals? Do speakers succeed in eliciting emotions they intend to? To answer those questions, we analyse real-time social media responses to selected televised debates, treating them as a marker of audience response. Then we employ methods from computational linguistics (Alswaidan and Menai 2020), i.e. machine learning models which identify sentiment and emotions expressed in language. This allows us to match the emotion appeals of the speakers with the audience reactions. Thus, the main contribution of this paper is the clear demarcation between speaker's and listener's perspective in pathos appeals, and joining them in one comprehensive model, along with the presentation of practical application of this model on the sample of natural language argumentation.

This paper is structured as follows: Sect. 2 presents the state of the art in three relevant areas—pathotic Argument Schemes, emotion-eliciting language, and emotion-expressing in natural language. Section 3.1 introduces material used in this paper: a

collection of arguments from pre-election debates together with social media reactions. Section 3.2 details our methodology for annotating and analysing pathotic Argument Schemes and emotion-eliciting language. Section 3.3 investigates audience emotional responses through automatic sentiment models applied to real-time social media data. Section 4 presents a quantitative summary of the observed pathos elements. Finally, Sects. 5 and 6 sum up emotional appeals and reactions, and point to limitations and perspectives of pathos analysis in natural language argumentation.

2 Literature Review

2.1 Emotional Appeals in Argumentation and Psychology

This section describes existing research in the two areas related to the two ways we can capture speaker's appeals to pathos: pathotic Argument Schemes and emotion-eliciting language.

2.1.1 Pathotic Argument Schemes in Modern Rhetoric

Following Aristotle's triad, researchers in argumentation have approached the concept of pathos from different perspectives over the years. Douglas Walton has devoted a significant number of his works to the role of various emotions in argumentation. His most fundamental claim, highly relevant to our work, is that emotions indeed do have a place in argumentation (Walton 1992). By adopting the descriptive, instead of the normative approach, Walton claims that emotions can be used argumentatively in persuasive dialogue. Instead of rejecting all emotional arguments as fallacious, an informal logician should carefully analyse the quality of argumentation, to observe whether the emotive component is relevant and not misleading. Selected emotions gained special attention from Walton, namely pity (Walton 1995) in the *argumentum ad misericordiam* and fear in the *Fear Appeal* argument (Walton 2013). By analysing natural language examples from legal and political discourse, Walton sheds light on the popularity and importance of fear-eliciting language as a persuasion device. Fear Appeal argument belongs to the type of *practical reasoning* argumentation schemes, i.e. schemes in which conclusion has the form of a call to action (Walton 2007a). In the approach adopted in this paper we follow Walton (1995) in accepting that pathos is part of everyday argumentation and we apply the concepts of Argument Schemes (Walton et al. 2008), allowing us for systematic observation of the use of emotion-eliciting schemes in political argumentation. Our focus is on schemes related to practical reasoning, as concluding with the call to action is expected to be found in political and commercial discourse (Walton 2010, 2007b).

The question whether emotional arguments are fallacious or not has been analysed by several researchers. For example Braet (1992) asks whether ethos and pathos could be argumentative in the same way as logos. In doing so, he recapitulates the discussion of scholars such as Eemeren (2018), who considers appealing to emotions as objectionable means of persuasion, and compares it with approaches

which acknowledge that pathos is present in everyday argumentation. Other researchers claim that the use of pathotic arguments can be considered legitimate if they are grounded in beliefs or cognitions that are reasonable (Manolescu 2006). This is reiterated in (Manolescu 2007), in which the author is following Walton's (2013) approach in which emotional appeals are not inherently fallacious.

In the study presented here we are using the method of manual annotation, i.e. marking of instances of Argument Schemes in natural language by human coders (annotators). In doing so, we follow e.g. Lindahl et al. (2019), who attempt at using argumentation schemes for annotation of Swedish political debate. The results of that study show significant differences in individual annotators' behaviour (resulting in low inter-annotator agreement). Furthermore, raters annotated a different number of Waltonian Argument Schemes, with selected schemes being more frequent than others. Similar results are reported in (Visser et al. 2021), which attempts at recognising Waltonian schemes and other argument types in natural language. Koszowy et al. (2022) reports the iterative process in which human annotation of rhetoric elements can be improved. For this reason, we propose our own simplified version of annotation guidelines moulded for the purpose of our study, focusing on argumentation schemes containing both call to action (practical reasoning) and the pathotic component (appeal to emotion).

Several researchers tackled the problem of emotions in argumentation from the empirical perspective, employing methods from both discourse analysis and psychological as well as physiological tools. Discourse analysis studies focus on emotion names occurring in natural language, referring to them as "said" emotions (Plantin 2019; Cigada 2019; Greco et al. 2022) in the context of persuasive dialogue. Certain authors add that emotions can be not only "said", but also "shown" or "argued" (Micheli 2010), that is they are expressed not only by direct emotion names but also other linguistic devices. This approach is adopted by Herman and Serafis (2019) who analyse how emotions can make certain argumentative moves more salient. Van Haaften (2019) connects the concept of emotions to strategic manoeuvring, analysing natural language argumentation in parliamentary speeches, which are also studied in terms of the use of metaphors for emotional appeals (Santibáñez 2010). In more empirical approach, (Cabrio and Villata 2018) presents emotion annotation in natural language arguments. We follow the approach used by those researchers in our focus on emotions in language, adding however the clear distinction between emotion-eliciting language and emotion-expressing language.

In natural language argumentation, Inference Anchoring Theory (IAT) models both argument structure and dialogical layer (Budzynska and Reed 2011a, 2011b), stemming from the tradition of philosophical approaches to linguistic pragmatics and speech act theory of Austin (1975), and Searle (1979). This theoretical approach allows us to capture premise-conclusion relations occurring in natural language. Thanks to the use of Argument Interchange Format (AIF) Database platform (Lawrence et al. 2012) we are able to construct computational models of natural argument structures. IAT and AIF have been previously used to analyse public debates and reactions in social media (Konat et al. 2016; Visser et al. 2020). Researchers have been pointing to the usefulness of the use of methods such as sentiment analysis for computational models of argumentation (Stede 2020). We follow this path by

applying two methods for analysis of appeals to emotions in argumentation structured in IAT: manual annotation of Waltonian schemes related to pathos, and automatic analysis of emotion-eliciting language with the use of lexicons stemming from psychological tradition.

2.1.2 Emotion-Eliciting Language and Psychological Lexicons

In psychology, particularly within the domain of cognitive psychology, a given word can be treated as a stimulus, comparable to many other stimuli types (sounds, images, videos) which will elicit emotional response in the listener. In this section we present how selected words used in argumentation can constitute emotion-eliciting language, and how we can apply existing psychological lists of such words (lexicons) to automatically identify their use in natural language argumentation.

There are two major approaches to studying the nature of emotions in cognitive psychology: dimensional and categorical. We chose to focus on these approaches as they form the basis for various lexicons and sentiment analysis tools still in use today (Mohammad and Turney 2010; Wierzba et al. 2021). The dimensional approach defines an emotional state by its placement in a multi-dimensional space (Russell 1980). Two or three such dimensions are employed most often, i.e. valence (positive to negative), arousal (low to high), and dominance (weak/submissive to strong/dominant). On the other hand, categorical approaches classify emotional experience into discrete categories. Ekman and Plutchik proposed two of such models that are commonly employed both in psychology and computational linguistics. Ekman's theory of basic emotions distinguishes 5 categories of emotions based on his research on facial expressions: anger, fear, sadness, disgust and joy. These are claimed to be universal, innate and hardwired (Ekman 1999). Plutchik's conceptualisation is also a biologically-based model; however, it identifies eight primary emotions. Five of these—anger, fear, sadness, disgust, and joy—are the same as those in Ekman's model, and Plutchik adds three additional emotions: surprise, trust, and anticipation (Plutchik 2003).

Emotion is a complex chain of events, starting from a cognitive evaluation of stimulus events that act as primary triggers. Evaluated information is transferred into actions which allows an individual to cope with the stimulus. This is accompanied by a feeling state (emotional experience), commonly referred to as the emotion, and followed by the reestablishment of an equilibrium state of an individual (Plutchik 2001). Emotional reactions are responses to biologically important stimuli, which enable organisms to prepare and respond to such stimuli. Emotion categories correspond here to different behavioural and physiological response patterns. Table 1 delineates Plutchik's idea of emotions as chains of events—each category of emotions is described by its characteristic stimulus event, cognitive evaluation of the stimulus, the feeling state, manifested behaviour, and effect.

A standard procedure to study emotional experiences in natural language involves the usage of affective lexicons. Such lexicons comprise words characterised in terms of several emotional attributes in accordance with a dimensional and/or categorical model of emotions (Mohammad and Turney 2010; Wierzba et al. 2021). Development of those linguistic resources involves studies with a large number of

Table 1 Plutchik's conceptualisation of emotions with characteristic events associated with them

Emotion Category (feeling state)	Stimulus event	Cognition	Behaviour	Effect
Joy	Gain of valued object	Possess	Retain	Gain resources
Fear	Threat	Danger	Escape	Safety
Anger	Obstacle	Enemy	Attack	Destroy obstacle
Sadness	Loss of valued object	Abandonment	Cry	Reattach to lost object
Disgust	Unpalatable object	Poison	Vomit	Eject poison

participants that are asked to rate a set of words on affective dimensions based on their individualised experiences (e.g., *how strongly a given word is associated with negative emotions?* on a 1–9 scale). Then, aggregation techniques (majority voting or averaging, for example) are employed to obtain the final ratings of such emotion-eliciting words that can function as stimuli in emotion research, similarly to emotion-evoking images (Saganowski et al. 2022). Processing of emotion-eliciting content enhances cortical response of information processing including semantic access, attention, and memory (Kissler et al. 2006). In our study we build on that body of research, by using existing psychological lexicons.

Emotions in argumentation can also be studied with experimental tools. Using physiological methods, Villata et al. (2017) analysed facial expressions of participants of simulated online debate, to capture their emotions. Researchers noted that with the increased amount of face expressions related to sadness, the number of arguments used by the debaters decreased, indicating the negative influence of this emotional state on the debate. This study indicates the importance of analysis of emotions as a tool for better understanding of public debate. In contrast to methods that employ physiological measurements, such as the analysis of facial expressions in the work of Villata et al. (2017), our study leverages the linguistic expressions of emotions as captured in text to examine the emotional dynamics within online debates.

2.2 Emotional Reactions and Sentiment Analysis

This section presents the overview of sentiment analysis methods from computational linguistics which allow us to analyse emotion-expressing language used by the audience in reaction to pathos appeals. Sentiment analysis methods are mostly based on the dimensional model of emotion (i.e. from positive to negative), supported by the assumption that the emotional state can be expressed in language.

State-of-the-art performance in automatic emotion recognition in text is systematically achieved by large pre-trained language models such as BERT (Devlin et al. 2019) as well as recurrent neural networks (RNN) (Alswaidan and Menai 2020). The latter technique is particularly successful in regard to classification in short text data such as social media reactions (Wang et al. 2016). Best performing teams of the SemEval-2019 task on (4-categorical) emotion detection report micro-averaged F1 score as high as 0.796 using transfer learning on BERT, long

short term memory (LSTM) networks or a combination of both (Chatterjee et al. 2019). However, the method is not free of limitations as Gordon et al. (2021) show that the evaluation of technical performance is not the best measure to evaluate the outcomes of machine learning (ML) systems in practice against human annotation.

Only recently, researchers foregrounded the distinction between the speaker's and the listener's perspectives in the data annotation and emotion analysis, which is crucial in the proposed model of pathos (Yang et al. 2009; Tang and Chen 2011; Buechel and Hahn 2017). Buechel and Hahn (2017) introduce EMOBANK corpus that implements this bi-perspectival approach, i.e. distinguishing emotions expressed by writers from emotions elicited in readers. In addition, the authors note that fine-grained modelling of emotions in computational linguistics "lack[s] appropriate resources, namely shifting towards psychologically more adequate models of emotion (...) and distinguishing between writer's vs. reader's perspective on emotion ascription" (Buechel and Hahn 2017, p. 578). Liu et al. (2013) study the relation between the comment writers' and the news readers' emotions. And Alsaedi et al. (2022) make use of comments corresponding to original social media posts as an additional source of information in emotion mining systems.

Tang and Chen (2011) in turn develop SVM-based classifiers for mining sentiment from the writer, the reader, and the combined writer and reader perspectives. The authors use conversations from an online micro-blogging platform, where both a poster and a replier could tag their own content with the experienced emotion (positive vs. negative) through the use of emoticons. A similar approach was adopted in (Berenguères and Castro 2017), where the authors investigate how being either a reader or a writer influences the perception of an emoji's sentiment.

Tang and Chen (2012) point out that the focus is on the writer's perspective in sentiment analysis research (i.e., emotions expressed in text). In contrast, the authors model writer-reader emotion transition in microblog posts. Thus, (Tang and Chen 2012) closely relates to our study. However, in the present work we propose a model that incorporates the analysis of the type of emotions the speaker is trying to elicit in the audience and their pathotic (emotional) response to it. Moreover, we study two different techniques speakers employ for appealing to pathos—the use of pathotic Argument Schemes and emotion-eliciting language. Appealing to emotions with Argument Schemes is a well documented phenomenon in argumentation theory (Walton 2013; Walton et al. 2008) which, however, has not caught the attention of researchers in computational linguistics yet.

There are also several works analysing reactions to politicians' statements in presidential debates. For instance, Diakopoulos and Shamma (2010) make use of Twitter data to measure sentiment expressed towards candidates in the first US Presidential debate in 2008 between Barack Obama and John McCain. However, to the best of our knowledge, we are the first to account for both perspectives in the model of pathos—the speakers' attempt to elicit emotions by the use of argumentation schemes and emotion-eliciting words, and the audience's pathotic response assessed in terms of emotions expressed in their reactions.

3 Material and Methods

3.1 Material

3.1.1 Natural Language Argumentation Material for Studying Emotional Appeals

To collect a large amount of diverse natural argument instances we decided to use the material of televised pre-election debates, in which the speakers are attempting at persuading the audience to vote for a selected party or candidate. We use both Polish and U.S.A. debates in order to compare the universality of pathos appeals in two languages and two cultures.

3.1.1.1 Pre-election Debates We analyse natural language argumentation from four pre-election debates, summary of which is presented in Table 2. In regard to the analysis of pre-election debates, we selected only the argumentative part of the debates, i.e. from the whole transcript we selected already existing instances of logos, marked using Inference Anchoring Theory (IAT).¹ Then, we labelled relations between premise-conclusion pairs with the proposed annotation scheme of pathotic arguments (presented in Fig. 6 in Appendix). Consequently, we restrict our analysis of emotion-eliciting language to the extracted argument structures. Here, we assume argument structures to have persuasive power and, therefore, to be the key element of persuasion.

We chose to investigate political debates from the United States and Poland for several key reasons. American political debates have received substantial scholarly attention and have a wealth of existing literature for reference. Furthermore, the US2016 corpus is both publicly available and thoroughly analysed (Visser et al. 2020, 2021), offering a solid foundation for our research. On the other hand, Polish political debates, while being televised and conducted, are less frequently the subject of academic enquiry. The Polish language is also sufficiently large to possess its lexicons of emotional words and machine learning models, making it a compelling case for investigation. Importantly, we had access to annotators proficient in Polish as their first language and conversant in academic English, thereby enabling high-quality annotations. It should be noted that our focus is not on making cross-cultural comparisons or drawing final conclusions based on these two datasets. Instead, we aim to observe overarching trends in emotional argumentation styles that manifest across both languages and diverse speakers.

3.1.2 Natural Language Material for Studying Audience's Reaction

To capture the audience's emotional response in terms of emotion-expressing language, we use the material from social media—live reactions to the televised debates described in the previous section. The audience of the debates is listening

¹ From corpora available on AIFdb: Infrastructure for the Argument Web (<https://corpora.aifdb.org/>)

Table 2 Pre-election debates data

Name	Date	Source	Authors	Size in words
Primary republican debate [D1-EN]	06.08.2015	https://corpora.aifdb.org/US2016R1tv	Visser et al. (2020)	16,382
First general election debate [D2-EN]	26.09.2016	https://corpora.aifdb.org/US2016G1tv	Visser et al. (2020)	16,204
Presidential debate May 2020 [D1-PL]	06.05.2020	https://corpora.aifdb.org/debatetvpmay20	Student annotators	10,803
Presidential debate June 2020 [D2-PL]	17.06.2020	https://corpora.aifdb.org/debatetvpjune20	Student annotators	11,750

Table 3 Social media data

Name	Date	Source	Authors	Size in words
SM1-EN-Reddit	06.08.2015	https://corpora.aifdb.org/US2016R1reddit	Visser et al. (2020)	10,765
SM2-EN-Reddit	26.09.2016	http://corpora.aifdb.org/US2016G1reddit	Visser et al. (2020)	11,484
SM1-PL-Twitter	06.05.2020	Unpublished	The authors of the article	150,504
SM2-PL-Twitter	17.06.2020	Unpublished	The authors of the article	270,755

to the arguments presented on the screen and, at the same time, is expressing their emotions using social media channels such as X (prev. Twitter) and Reddit.

3.1.2.1 Social Media Reactions We made use of corpora comprising Reddit comments related to the US primary debates and available on the AIFdb platform—that is, US2016R1reddit and US2016G1reddit corpora (Visser et al. 2020). In order to measure pathotic response in Polish data, we collected comments written on X (prev. Twitter) during the time of live presidential debates. Specifically, we gathered tweets that contain a proper hashtag related to the debate,². Data for the analysis of expressed emotions is summarised in Table 3.

In order to observe pathotic response, we matched the arguments from debates with social media reactions and for each time unit we computed proportions of comments that express positive and negative sentiment, and each category of emotions.

In regard to Polish data, we combined arguments from debates with social media response based on time (1 min units, i.e. a time span dedicated for a single politician's speech in each round of the debate). In addition, we applied time correction to this set of data—manual inspection of data showed 1 min delay in reaction on social media compared to live political debate.

Regarding English data, in order to match arguments with reactions on Reddit we made use of inter-textual correspondence annotation from the US2016itc corpus (Visser et al. 2020). The reason is that we did not have access to timestamps of posting social media comments. Thus, we relied on the annotation from the US2016itc corpus, where politicians' arguments were manually paired with selected Reddit comments. We retrieved all argument-comment pairs from the US2016itc corpus related to the first Republican and the first General election debates. Then, we expanded the collected dataset by adding up to 10 comments from US2016R1reddit and US2016G1reddit corpora surrounding those manually matched Reddit comments from the US2016itc argument-comment pairs (Visser et al. 2020). In other words, a politician's claim from debate is manually matched with a comment from Reddit, which is a direct response to the politician's claim. In US2016R1reddit and

² The following hashtags were used: “debata”, “czasdecyzji” (En. *debate decision time*).

US2016G1reddit corpora there are more comments related to the first Republican and the first General debates from which we retrieved up to 10 comments that follow each manually-matched comment in the US2016itc corpus.

Basic pre-processing was applied to social media data. It includes removal of blank tweets, tweets with only URL content and duplicated tweets. Textual content was also normalised by conversion to lowercase, replacement of user mentions (@) with a “@user” token, and removal of new lines and extra spaces.³

3.2 Recognition of Emotional Appeals in Natural Language

Two methodologies are applied to the argumentation of pre-election debates in order to capture the ways speakers can perform pathos appeals. First, we conduct manual annotation with the proposed taxonomy of eight pathotic Argument Schemes. Second, we perform automatic identification of emotion-eliciting language using psychological lexicons.

3.2.1 Annotation of Pathotic Argument Schemes

We employ manual annotation to categorise argument structures into specific instances of pathos-related Argument Schemes. We hypothesise that instances of arguments corresponding to these schemes can be reconstructed in natural language argumentation from pre-election debates. For this purpose, we adopt the taxonomy of Argument Schemes by Walton et al. (2008), applied to natural language argumentation on a material of political debates. We simplify the guidelines regarding recognition of instances of pathos-related Argument Schemes and focus on decoding the speaker’s intention of appealing to emotions in order to increase the persuasive force of her argumentation.

The selection of the eight Argument Schemes as pathos-related is our own categorisation, not that of Walton, Reed, and Macagno. We have chosen these schemes based on criteria that include either a named emotion by the authors (e.g., fear), or an intention of valuation mentioned in the scheme (e.g., “bad consequences” in the Argument from Threat). In the case of the Argument from Positive Consequences and Argument from Negative Consequences, our understanding is influenced by previously introduced Plutchik’s model of emotions, where emotion starts with a cognitive evaluation (either as positive or negative) of stimulus events that act as primary triggers. In this sense, the “Positive/Negative Consequences” schemes align with our conceptualisation of emotions. We regard eight Argument Schemes as pathos-related:

1. Argument from Positive Consequences (APC),
2. Argument from Negative Consequences (ANC),
3. Argument from Fear Appeal (AFA),

³ All code used for analysis in this paper is available on GitHub: <https://github.com/barbara-k/pathos>.

4. Argument from Danger Appeal (ADA),
5. Argument from Threat (AT),
6. Slippery Slope Argument (SSA),
7. Argument from Waste (AW),
8. Argument from Need for Help (ANH).

The primary purpose of political speeches, which are part of pre-election political campaigns, is to convince the public to vote for a given politician or a political option represented by her (Hinton and Budzyńska-Daca, 2019). Therefore, in the construction of our annotation scheme we follow the assumption of practical reasoning being present in politicians' arguments. In the broader context of a debate, each argument contributes to the general practical reasoning process with the main, enthymematic conclusion "I should be the president" or "Vote for me", although it is not always explicitly expressed by speakers themselves. As a result, the proposed annotation scheme is not suitable to the same degree in every genre. It should be applicable to most genres of political debates (e.g. pre-election, citizen dialogues, political speeches, consultations), where a call to action is an underlying principle.

The annotation scheme proposed in the current work is based on the critical work by Walton et al. (2008). However, in our operationalisation (see Appendix 2) we have revised the argumentation schemes to concentrate on the central emotional intent ("the message of the argument" in our annotation scheme). By this, we mean the primary emotional tone or effect that the speaker aims to evoke, which is integral for this argument. Upon identifying whether an emotional intent is central to the argument, we further categorise this into its "essence", i.e. dominant emotional characteristic, be it predominantly positive, negative, or oriented towards eliciting a specific emotion like fear. This allows us to better understand the argument's core emotional dimension. On a cognitive level, the speakers attempt to induce a particular emotional state in the audience in order to influence their perception and processing of information, and as a result direct their behaviour. The speakers do so, on a language level, using linguistic means of pathos-related Argument Schemes. We design our annotation scheme to aid our annotators in recognising a particular instance of such pathos appeals.

To capture pathos appeals performed by the speakers on the level of selected Argument Schemes this method was applied to four debates presented in Sect. 3.1. A team of 5 student annotators was recruited for this purpose. The first language of all annotators was Polish, they also had a minimum B2 level of English language and used it during academic courses. They underwent a short training session conducted by one of the authors of the current work. Then, each of the annotators was assigned to annotate 78 argument maps⁴: arguments from 50 maps were annotated by all 5 annotators (in order to calculate inter-annotator agreement) plus 28 unique maps that were randomly assigned to each annotator. Argument maps were annotated using the Online Visualisation of Argumentation (OVA+) tool (Janier et al. 2014). As mentioned in Sect. 3.1, annotators worked on existing OVA+ maps stored

⁴ The term "argument map" refers to an excerpt of analysed text, usually 500–1500 words, a unit used by OVA+ (Janier et al. 2014) annotation tool and AIFdb (Lawrence et al. 2012).

in AIFdb where *logos*, i.e. premise-conclusion structure, was already marked by previous research teams.

In the annotation of pathos-related Argument Schemes, raters follow the annotation guidelines illustrated in Appendix in Fig. 6. Annotators who perceive an argument as pathos appealing, should first answer the question included in the top frame of the annotation scheme: “Does the speaker intend to elicit emotions?”. If the answer to this question is “No”, then the annotator should not annotate this argument as pathos-related; if the answer is “Yes”, then the annotator should move to the next step. To answer this question the annotator should follow their intuition, but also look for linguistic cues present in argument structures: emotional words, phrases, metaphors, and other rhetorical figures. For instance, those cues comprise emotional words such as “children”, “war”, “terror”, and emotional metaphors, e.g. saying that something is “heartbreaking”.

In the next step the annotator should consider whether the argument is based on the causal relation. The argument based on a causal relation should contain the information about an event A, which will lead to an event B. Then the annotator should consider whether the event B is bad and unwanted from the listener’s perspective. If so, the annotator should move to the proper frame and annotate this argument as either an Argument from Negative Consequences (ANC) or one of its subtypes, following the information provided in the Subtypes of ANC section presented below. If the consequences mentioned in the argument are not bad and unwanted from the listener’s perspective, then the annotator should consider if the consequences mentioned in the argument are positive from the listener’s perspective. If yes, then the annotator should annotate this argument as an Argument from Positive Consequences.

The annotator should further investigate whether the argument that is considered an Argument from Negative Consequences contains information included in the Subtypes of ANC list (presented below). An argument can be classified as a certain subtype of an Argument from Negative Consequences if it contains the specific information. One argument cannot be labelled as fulfilling requirements of more than one Argument Scheme. However, when the argument cannot be classified as one of the subtypes of arguments from the Subtypes of ANC list, but is recognised as a causal argument, in which bad and unwanted consequences are mentioned, then it should be annotated as an Argument from Negative Consequences. Key information that should be mentioned if the argument is classified as a subtype of an Argument from Negative Consequences (ANC) is as follows:

1. Argument from Fear Appeal
 - (a) The message of the argument should be to bring about the only way to prevent an unwanted event.
2. Argument from Danger appeal
 - (a) The message of the argument should be not to bring about the event that will cause danger (unwanted event).
3. Argument from Threat
 - (b) The argument should express threat: the speaker should be in the position to bring about the unwanted (from the listener perspective) event, and it should

be mentioned that the speaker will execute a threat if the listener will do what the speaker does not want the listener to do.

4. Slippery Slope Argument

- (a) The message of the argument should be that one event will cause a chain of consequences that will lead to the unwanted and disastrous event.

If the argument is not based on causal relation, then the annotator should follow the instructions located on the right side of the annotation scheme (see Fig. 6 in Appendix). The annotator should answer the question whether the speaker in this argument is referring to the personal attachment or the personal concern of the listener. If no, then that argument should not be annotated; if yes, then the annotator should check if the argument contains information about wasted effort. If so, then it is an Argument from Waste. However, if the answer is no, the annotator should inspect if the speaker in this argument is calling for help by the induction of empathy. If the argument contains such information, it should be annotated as an Argument from Need for Help, otherwise the annotator should not annotate this argument.

Following Koszowy et al. (2022), we conducted a second iteration of annotation. This time a team of annotators comprised 3 individuals from the first iteration. Results of the second iteration of manual annotation of Argument Schemes remain stable—we obtain similar levels of inter-annotator agreement (IAA) coefficients. Thus, we assume that the proposed operationalisation of Waltonian Argument Schemes (Walton et al. 2008) is suitable for the recognition of pathotic Argument Schemes in natural language argumentation.

3.2.2 Identification of Emotion-Eliciting Language

To observe the second way the speaker can elicit emotions in the audience, i.e. by using emotion-eliciting words, we used an automated method of searching for emotion-eliciting words, based on lists created by psychologists (Mohammad and Turney 2010; Wierzba et al. 2015, 2021). This method was applied to all four debates presented in Sect. 3.1.

First, each conclusion-premise pair was lemmatised with the use of SpaCy library. Second, we retrieved emotion-eliciting words with the use of selected affective lexicons. In regard to Polish language, we chose Emotion Meanings (Wierzba et al. 2021) and Nencki Affective Word List (NAWL) (Wierzba et al. 2015). In the case of the first lexicon which comprises 6000 word meanings, we made use of ratings for Ekman's 5 primary emotions—anger, fear, sadness, joy, and disgust. Similarly for NAWL, we considered all 2902 words assessed in terms of Ekman's 5 basic emotions. We applied scale normalisation to emotion ratings to unify them into one lexicon, harmonising differing original scales to a 0 to 1 range. As a result, we obtained an emotion lexicon comprised of ratings for almost 8000 words. Regarding English language, we selected the NRC Word-Emotion Association Lexicon (EmoLex) with 5961 terms assessed in terms of primary emotions (Mohammad and Turney 2010). Similarly, we chose the same 5 basic emotions and applied scale normalisation to values in the EmoLex.

Third, the intensity of categorical emotion-eliciting language was computed for each argument unit according to the procedure described in Algorithm 1 (see Appendix 1). That is, we computed the average intensity of emotional appeals in a given text based on retrieved emotion-eliciting words listed in psychological lexicons, where categorical emotion is provided (e.g. joy, sadness, fear). Those values were then used in the correlation analysis between emotion-eliciting language (on the side of the speaker) and emotion-expressing language (on the side of the audience).

3.3 Using Sentiment Analysis to Capture Audience Reactions

The audience of pathos appeals in pre-election debates can use live social media channels to describe their emotional reactions. To analyse the presence and intensity of emotion-expressing language in social media reactions to the debates, we employed automatic machine learning models for sentiment analysis adopted from computational linguistics.

We made use of Tw-XLM-roBERTa-base model pre-trained on a multilingual tweets dataset and fine-tuned for sentiment recognition (Barbieri et al. 2021). It is made publicly available by the authors in the Transformers library (Wolf et al. 2020).⁵ This model allows to classify sentiment expressed in text into one of the following categories: negative, neutral, and positive. It was employed for the recognition of sentiment in English social media comments. With respect to Polish language data we developed the PaRes model. That is, we additionally trained the Tw-XLM-roBERTa-base model on a 1000 Polish data sample manually annotated with sentiment by one of the authors of the present article.

Because of the scarcity of resources for Polish language, we decided to develop our own - PaREMO - model for the recognition of emotions expressed in the audience's reactions. For this purpose, we made use of data created by other researchers and available for the scientific community. Thus, training data for our model comprises the following datasets: CARER (Saravia et al. 2018), GoEmotions (Demszky et al. 2020), and SemEval 2018 subtasks 1 and 5 (Mohammad et al. 2018). Each of them was manually annotated with basic emotions. For the purpose of the study, we chose the following categories of emotion: anger, fear, joy, surprise, sadness, disgust, and neutral. Collected corpus comprises over 58,000 samples – 85% was used for training the model and the remaining 15% comprised a test set.

As a text representation method we employ LASER (Language-Agnostic Sentence Representations) multilingual sentence embeddings developed by Facebook (Artetxe and Schwenk 2019). Language-agnostic means that sentences written in different languages that convey the same semantic meaning are mapped to the same place in a multi-dimensional space. Therefore, this technique enables us to train our deep learning model on English data (because of scarcity of resources in Polish) and subsequently detect emotions in Polish language data.

⁵ <https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment>

Table 4 Summary of models employed for sentiment and emotion recognition in English and Polish data

Model name	Task	Language	Macro-F1	Micro-F1
Tw-XLM-roBERTa-base	Sentiment	English	0.635	0.676
PaRes	Sentiment	Polish	0.696	0.720
PaREMO	Emotion	English and Polish	0.498	0.594

Summary description and performance results for all 3 models used in the study are presented in Table 4. An evaluation against manual annotation for PaRes and PaREMO models was conducted on a 15% subset of their respective training data. In regard to the Tw-XLM-roBERTa-base model we run validation on a 1000 sample of English social media comments related to a political discourse and annotated with sentiment by one of the authors.

We run two types of correlation analyses on the material from debates and social media. First, we examine the relation between the usage of pathotic arguments by speakers in debates and emotions expressed by the audience on social media. Second, we investigate the association between the speakers' emotion-eliciting language and the audience's emotion-expressing language on social media.

4 Results

4.1 Pathos in Natural Language

4.1.1 The Use of Pathotic Argument Schemes

As a result of the annotation process we obtained 190 maps (with 1621 arguments, and 3774 individual annotations in total) from 4 debates corpora—2 Polish and 2 US. They were annotated by a team of 5 raters. A sample of 50 maps with 529 arguments was annotated by all raters in order to compute inter-annotator agreement (IAA) metrics. Summary of annotated argument schemes is presented in Table 5.

In Fig. 3 we present results of the manual annotation study. Figure 3 a. shows that emotional appeals constitute 52% (849 instances) of all the arguments in the analysed debates, whereas non-pathotic arguments represent 48.0% (772) of the arguments in the final corpora, which means that they were considered by annotators as not appealing to pathos. The summary presented in Fig. 3b. indicates that the two most frequent arguments related to pathos are arguments from Positive Consequences (340 arguments; 21.0%) and arguments from Negative Consequences (299; 18.4%). Arguments from Need for Help, Argument from Danger Appeal, and Argument from Fear Appeal constitute respectively 5.0% (81), 4.3% (69), and 2.7% (44) of the re-annotated corpora. Slippery Slope arguments are present in 0.7% (12) of the sample and Argument from Waste was assigned to 0.2% (4) of the arguments. Argument from Threat was not annotated by the raters in any case.

Table 5 Results of the argument schemes manual annotation study

Argument scheme	No	Percentage
Non pathotic arguments	772	47.6
Positive consequences	340	21.0
Negative consequences	299	18.4
Need for help	81	5.0
Danger appeal	69	4.3
Fear appeal	44	2.7
Causal slippery slope	12	0.7
Waste	4	0.2

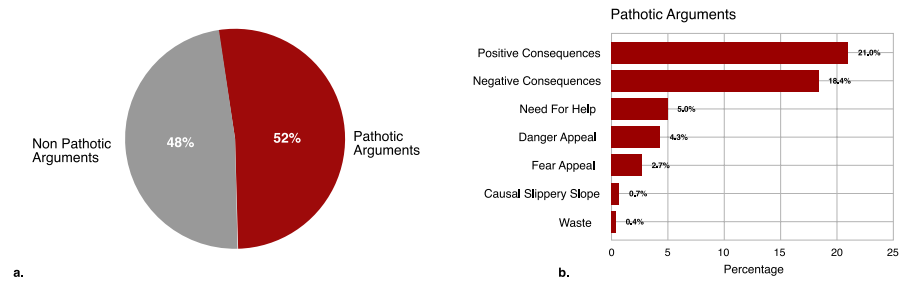


Fig. 3 Pathotic schemes distribution in Argument Schemes manual annotation study. Argument schemes related to pathos are marked by red colour

In Table 6 we present the results of inter-annotator agreement. Agreement could be interpreted as fair in the case of 4 out of 5 Argument Schemes (Fleiss $\kappa = 0.3-0.4$). In regard to the annotation of Arguments from Negative Consequences, we observe moderate agreement ($\kappa = 0.6$). However, some Argument Schemes (Danger Appeal, Arguments from Waste, Causal Slippery Slope) were too infrequent in a sample to estimate reliable kappa coefficients. This infrequency causes the 'prevalence problem' (Eugenio and Glass 2004), where κ reflects agreement mostly for prevalent classes, but is less reliable for infrequent ones. Despite this, percent agreement was above 0.9, indicating high IAA overall, but necessitating cautious interpretation.

In addition, we observe perfect agreement in 44% of the annotated sample given 3 best annotators. Specifically, all raters agree in: 2 cases on Fear Appeal, 3 cases of Need for Help, 49 arguments from Positive Consequences, 63 cases of arguments from Negative Consequences, and in 116 examples of Non Pathotic Arguments annotation.

The analysis of natural language argumentation in terms of searching for instances of pathos-related Argument Schemes reveals several interesting patterns. First, arguments appealing to emotions constitute more than a half of all arguments found in pre-election debates. Even if we take into account the possibility of cognitive bias of our annotation team, which was conditioned to specifically search for this type of arguments, this result suggests that pathos-related

Table 6 Results of the inter-annotator agreement analysis given 3 (out of 5) best annotators

Argument scheme	Fleiss' Kappa	Interpretation
Negative consequences	0.6	Moderate
Positive consequences	0.4	Fair
Non pathotic arguments	0.4	Fair
Need for help	0.4	Fair
Fear appeal	0.3	Fair

Other schemes were too infrequent to estimate reliable Kappa coefficients

schemes deserve further attention and study, especially in terms of their effectiveness. Such a frequent use by trained politicians may suggest at least a strong belief that this type of argumentation is highly convincing. Assessing the persuasiveness of such arguments goes beyond the scope of this study, as it is centred on structured observation, it does not control for dependent variables, thereby limiting our ability to assess the actual persuasive effectiveness of emotional appeals. We can identify phenomena in the corpus, but these observations remain suggestive of trends rather than definitive evidence of effectiveness. A controlled experimental approach, such as the methodology proposed by Villata et al. (2017), would allow for the measurement of the effect size, hence providing more definitive evidence on the persuasiveness of appealing to emotions within argumentation.

After careful analysis of the results of inter-annotator agreement we decided to create the final corpora according to the following guidelines. Re-annotated corpora consists of 50 argument maps annotated with a majority label and 112 maps annotated by a single rater (5x28 argument maps). As a result of the re-annotation study we obtain 190 argument maps with 1621 arguments in total, annotated with pathos-related Argument Schemes. They have been assigned to 4 new corpora, which correspond to the source corpora. The names and the links to the new corpora with the annotation of pathotic arguments are presented in Table 10 in Sect. 3 of Appendix.

The second main conclusion comes from the “long tail” distribution of the schemes, which repeats the distributions found in similar annotation studies (Visser et al. 2021). With Negative and Positive Consequences schemes being the most frequent, we can assume some type of an artifact, where more general categories tend to be annotated more frequently and with higher agreement. Still, the results seem reasonable in the context of the specific type of discourse we have selected. In political debates, the speakers want to present practical reasoning where they will encourage the audience to either take upon an action or refrain from one, depending on the foreseen consequences. Fear and Danger Appeals are still quite frequent, albeit the “scare tactics” do not possess the special place in the distribution, as might have been assumed based on the legal argumentation analysis provided in Walton (2013). This might be again a case of the genre, where politicians do not want to appeal overtly to fear, however comparative systematic studies will be needed to answer this question fully.

For such studies, we would need stable and reliable annotation guidelines, allowing for the inter-annotator agreement to stabilise over the moderate level (i.e. 0.6 Fleiss's κ). As our results are showing, this task is difficult, and we identify two main reasons for that. First, Argument Schemes were not conceived as being ideally mutually exclusive categories. The models provided by Walton et al. (2008) almost always fit only partially into the cases of real-life argumentation instances. Further theoretical work is needed in the direction of creating operationalised annotation guidelines. We provide our proposal of such guidelines in Appendix 2, however more iterations of testing and improvement, beyond the two presented in this paper, are needed. Second difficulty encountered in annotating pathos-related Argument Schemes is related to the inherently highly subjective nature of emotions. When an annotator is asking themselves a question "Which emotion is the speaker attempting to elicit?" the answer may depend not only on the linguistic surface and argumentative structure of the text they have in front of them, but equally on the individual psychological predispositions, life history, cultural context etc. Even though we cannot take all these features into account, certain methods allow us to obtain some level of inter-subjective description of emotion appeals. This includes alternative measures, such as majority vote, but also new techniques coming from the field of AI, allowing to capture individual assessment. Here, Miłkowski et al. (2021) propose Personal Emotional Bias (PEB) metric as a measure of an individual's tendency to annotate different categories of emotion. In further studies it could be adopted to the annotation of pathos-related Argument Schemes to investigate those individual differences in the annotation of emotion-appealing arguments. Previous studies show emotional arguments are effective in persuading individuals with a certain personality profile. With the use of the Big Five model Lukin et al. (2017) demonstrate that conscientious, open, and agreeable people are convinced by emotional arguments rather than factual arguments. The recently proposed perspectivist approach (Kocoń et al. 2021; Abercrombie et al. 2022) is another area of interdisciplinary studies, combining social science and computer science, that is worth studying in future as the next step in understanding the subjective nature of emotion perception and persuasive argumentation. Another solution comes from the field of psychology, where researchers have attempted at representing generalised measurements of emotional reactions in terms of lexicons of emotional stimuli, i.e. emotion-eliciting words. For this reason, we decided to support our manual analysis of Argument Schemes with the automated analysis of stimuli words.

4.1.2 The Use of Emotion-Eliciting Language

All natural language arguments from pre-election debates were analysed in terms of the presence of emotion-eliciting language in them. We found that almost 95% of arguments contain emotion-eliciting words (Fig. 4). On average, there are 2.75 emotion-eliciting words in each argument (min = 0, max = 23). As a result, emotion-eliciting words comprise 12.67% of words in argument structures, on average (min = 0%, max = 60.0%). Most often there are 2 emotion-eliciting words in arguments

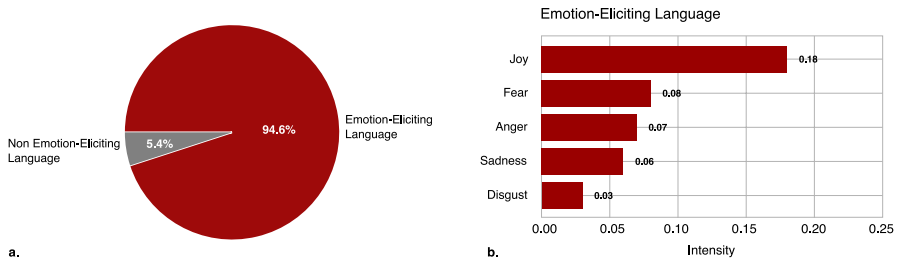


Fig. 4 Percentage of argument structures with emotion-eliciting language (a) and average intensity of five basic types of emotion-eliciting language in natural language argumentation (b)

(21.79% of cases)⁶. Interestingly, emotion-eliciting words in Polish debate data comprise 22.36% of words (in arguments). In US debates in turn it is only 8.52%. On average, each emotion-eliciting word in Polish data is repeated 4.49 times, and in regard to US data it is 5.60 times. With respect to US data 18.8% of tokens were recognised by the affective lexicon and 24.5% of tokens in Polish data. In addition, we find argument premises to be slightly more dense in emotion-eliciting words (on average) compared to argument conclusions –13.12% vs. 12.90% of words in those argument structures, respectively. Examples 2 and 3 present two of such arguments densely packed with emotion-eliciting words (depicted in bold font). In addition, we find joy to be the emotion most intensely appealed to, followed by fear, anger, sadness and disgust. Summary results are depicted in Fig. 4, which presents how over 94% of arguments contain some form of emotion-eliciting language (a), and how joy and fear are the most intense emotions in those appeals (b).

Example 2 Andrzej Duda (May 6, 2022; TVP):

Premise: Anyway, **adoption** by same-sex **couples** is **forbidden** by the Polish constitution⁷.

Conclusion: **Adoption** by same-sex **couples** is **absolutely** out of the question⁸.

Example 3 Władysław Kosiniak-Kamysz (June 17, 2022; TVP):

Premise: We shall appreciate **hard-working people** by raising **tax-free amount**⁹.

Conclusion: We should not **punish** entrepreneurs¹⁰.

The lexical method used in analysis of debates has multiple limitations previously recognised in literature (Alessia et al. 2015; Warriner and Kuperman 2015). First, it is context-insensitive as affective lexicons usually comprise a single dictionary form

⁶ Top 10 emotion-eliciting words in Polish data: ‘prezydent’, ‘rząd’, ‘państwo’, ‘rok’, ‘musieć’, ‘pan’, ‘móc’, ‘zdrowie’, ‘podatek’, ‘chcieć’. Top 10 emotion-eliciting words in US data: ‘good’, ‘bad’, ‘tax’, ‘money’, ‘government’, ‘military’, ‘deal’, ‘problem’, ‘leave’, ‘pay’.

⁷ Original: “Zresztą dzisiaj polska **konstytucja** **absolutnie** na to [adopcję dzieci przez pary jednopłciowe] nie **pozwała**.”

⁸ Original: “Jest to [adopcja dzieci przez pary jednopłciowe] **absolutnie** wykluczone.”

⁹ Original: “Doceńmy **ludzi ciężkiej pracy**, przez **kwotę wolną** od **podatku** wyższą.”

¹⁰ Original: “Nie wolno ich [przedsiębiorców] **karać**.”

of words instead of word meanings. Techniques such as word sense disambiguation could improve the accuracy of emotion analysis with lexicon-based methods to a certain degree (Jose and Chooralil 2015). Second, we are able to assess the intensity of emotion-eliciting language only at a certain confidence level as there is a finite number of words available in psychological affective dictionaries. Thus, we could account for most but not all emotion-eliciting words employed in debates with the use of this method.¹¹

The Pollyanna effect (Boucher and Osgood 1969)—a well documented phenomenon in regard to human evaluation of emotional stimuli—is another limiting factor, observed also in our data analysed with the use of lexical method. The Pollyanna effect, also called the positivity bias, is the result of the human tendency to use value-laden positive language more frequently. Taboada et al. (2011) argue that it is commonly observed in lexicon-based approaches to sentiment analysis, and as a result could degrade accuracy of emotion recognition tools.

4.2 Emotional Reactions

4.2.1 Reactions to Pathotic Argument Schemes

Summary statistics of expressed emotions recognised in the audience reactions on social media is presented in Table 7. Results indicate that the model employed for the task was biased towards one emotion—surprise. Therefore, we decided to discard instances recognised with the emotion of surprise from correlation analyses.

First, we test the relation between general emotion-eliciting language and emotion-expressing language in terms of the use of categorical emotions. The analysis is conducted separately for the US and Polish corpora. No statistically significant results are observed here. Therefore, we decided to run further (detailed) analyses, separately for each of the debate corpora, Argument Scheme and emotion category following related works (Villata et al. 2017). We hypothesise that different topics of discussion could have an impact on the strength and direction of associations between eliciting and expressed emotions.

Results of this detailed correlation analysis between the occurrence of arguments that appeal to pathos and emotions expressed by the audience on social media are presented in Table 8. For the purpose of this study we employed the point-biserial correlation which is used to measure the strength of association between two variables, when one variable is continuous and the other is dichotomous (binary).

In terms of specific categorical emotions, we would like to draw the reader's attention to one particular emotion, i.e. anger. Our findings indicate that politicians are successful in their attempts to induce negative emotions in the audience by appealing to negative emotions via the structure of argumentation (Argument Schemes). Figure 5 depicts the relation between politicians' arguments from negative consequences and anger expressed in social media reactions. Appeals to negative emotion are employed by speakers in order to scare the audience and induce

¹¹ In regard to Polish language 7.34% of words from the selected affective dictionary were recognised; regarding English it is 3.10%.

Table 7 Proportions of emotions and sentiment recognised in the audience response on social media

Expressed Emotion	Percentage
Surprise	57.6
Anger	18.0
Joy	11.7
Sadness	7.8
Fear	4.7
Disgust	0.3
Negative sentiment	58.0
Neutral sentiment	27.9
Positive sentiment	14.2

Table 8 Results of point-biserial (r_{pb}) correlation analysis between the presence of pathetic arguments in debates and expressed emotions in the audience' reactions on social media

Debate	Argument scheme	Expressed emotion	r_{pb}
D1-EN	Negative consequences	Negative sentiment	-0.55*
D2-PL	Negative consequences	Anger	0.86**

Only statistically significant results are depicted

* $p < 0.05$, ** $p < 0.01$

negative emotions in them. They are often employed by politicians in their campaigns in order to convince the audience to vote for them and not for the opponents. Moreover, appeals to negative emotions are used in every-day language and product advertisement (Walton 2013). Its persuasive force comes from the structure of this argument—a speaker states that harmful consequences will occur unless the listener takes actions advocated by the speaker. Therefore, the listeners are faced here with an ultimatum: they take the advocated action or the harmful consequences will occur.

We are not the first to report appeals to negative emotions as a common strategy employed in a political discourse. Maďarová (2015) found that fear mongering narratives were frequently employed by conservative groups against gender equality and the human rights of LGBT people in Slovakia. In turn Bourse (2019) points to the persuasive power of “loaded words” on the example of political speeches on drug reform in the United States. Negative emotion-eliciting words such as “tragically”, “pain” and “death” were the second most frequent category used in those speeches. Similarly, we find many cases of usage of emotion-eliciting words such as “terrorist”, “fight” and “revolution” in Arguments from Negative Consequences.

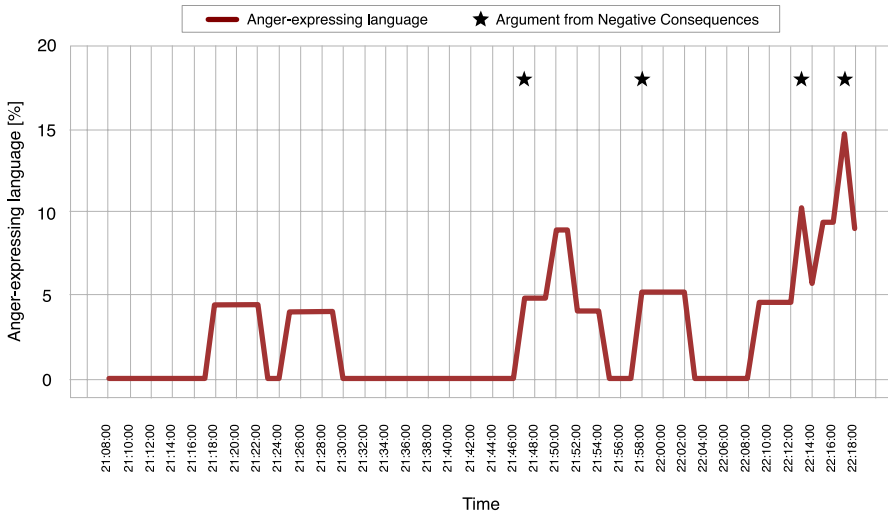


Fig. 5 The use of Arguments from Negative Consequences is followed by an increase in the usage of anger-expressing language by the audience. Asterisks reflect the use of Arguments from Negative Consequences in a debate. Line indicates the percentage of replies on X (prev. Twitter) that contain anger-expressing language. A moving average with a size of 5 min was employed to assess the change in use of anger-expressing language in audience's reactions

4.2.2 Reactions to Emotion-Eliciting Language

In regard to the second type of analysis, i.e. reactions to emotion-eliciting words, we obtain one statistically significant result regarding the relation between general saturation of emotion-eliciting and emotion-expressing language— $r = 0.42$, $p < 0.01$ in the case of the D1-EN debate. This means, the more emotion-eliciting words the speakers are using, the stronger emotional reaction from the audience. Similarly, we decided to conduct a further correlation analysis with respect to categorical emotions. Assessed proportions of expressed emotions were used for calculating Pearson's r correlation coefficients. In addition, we applied logarithmic transformation to the data in order to account for skewed distributions. Statistically significant coefficients are presented in Table 9.

We observe 9 statistically significant relations between emotion-eliciting language and emotion-expressing language. We observe one strong correlation between anger-eliciting words in argument premises and expressed sadness in the audience's response ($r = 0.77$). Three correlations can be interpreted as weak: anger-eliciting words in argument premises and expressed joy: $r = 0.27$; fear-eliciting words in argument premises and expressed joy: $r = -0.29$; sadness-eliciting words in argument conclusion and expressed fear: $r = 0.26$. The other coefficients are moderately strong: joy-eliciting words in argument conclusion and expressed disgust: $r = 0.49$; joy-eliciting words in argument conclusion and expressed anger: $r = -0.36$; disgust-eliciting words in argument conclusion and expressed joy: $r = -0.32$; joy-eliciting words in argument conclusion and expressed anger: $r = -0.32$; fear-eliciting words in argument structures and expressed disgust: $r = 0.48$.

Table 9 Pearson's r correlation coefficients between emotion-eliciting language in pre-election debates and emotion-expressing language in social media reactions

Debate	Emotion-eliciting language	Emotion-expressing language	Pearson's r
D1-EN	Premises anger	Sadness	0.77*
D1-PL	Conclusion joy	Disgust	0.49*
	Conclusion joy	Anger	-0.36**
	Conclusion disgust	Joy	-0.32*
	Premise anger	Joy	0.27*
	Premise fear	Joy	-0.29*
D2-PL	Conclusion joy	Anger	-0.32*
	Conclusion sadness	Fear	0.26*
	Full argument fear	Disgust	0.48*

* $p < 0.05$, ** $p < 0.01$

Elicitation and expression of negative emotions (anger, fear, sadness, disgust) tend to be positively correlated with each other, although the strength of association varies from weak to strong. The weak association could be observed for sadness-eliciting words in argument conclusion and the expression of fear on social media. A moderate correlation is observed between the use of fear-eliciting words in politicians' arguments and disgust expressed by the audience. Categories of positive and negative emotions tend to be negatively correlated with each other as in the case of joy and anger in Polish corpora. In addition, we could observe that the relation between those two emotions is stable across corpora: $r = -0.36$ and $r = -0.32$ in the case of D1-PL and D2-PL corpus, respectively. However, we also observe a positive association between joy-eliciting words in argument conclusions and disgust expressed by the audience, and anger-eliciting words in argument premises and expressed joy in the D1-PL corpus. No statistically significant correlation was found for the D2-EN corpus. Based on results from Table 9 one could conclude that "lexical" appeals to emotions are infelicitous. Results of our study indicate that the audience tend to respond with emotions that are different from those that politicians attempted to elicit. In the literature one could find similar studies with the use of visual stimuli that corroborate our findings. For instance, Saganowski et al. (2022) intended to invoke particular emotions by pre-designed film clips, however, participants reported experiencing not only the intended emotions (i. e., those that films were designed to invoke) but also other categories. For instance, in the case of eliciting anger participants reported feeling anger, disgust, fear, sadness and surprise, and eliciting fear induced fear, disgust, sadness and surprise. Therefore, our results are in line with findings in psychology that certain emotional stimuli tend to evoke many different emotional states, not only the intended ones.

5 Discussion

Using the case of pre-election debates and online reactions to them, we have presented two ways speakers can appeal to the emotion of the audience: by using pathetic Argument Schemes and with emotion-eliciting language. To capture the audience reactions, we applied the third method - sentiment analysis for expressed emotions. We presented the model and method of analysing appeals to emotions in natural language using an interdisciplinary approach. From argumentation theory we prepared our own adaptation of Argument Schemes relating to pathos, and we applied it to the large sample of natural language argumentation from Polish and English. From psychology we incorporated lexicons of emotion-eliciting words, and searched for their presence within argument structures (i.e. premises-conclusion pairs). In order to match the appeals with possible reactions, we collected real-time social media reactions to the debates and applied machine-learning models from computational linguistics, to observe emotion expressed in language.

First, the results of our Argument Schemes study indicate that pathos-related instances constitute over half of the natural language arguments collected in the corpora (52%). Therefore, the usage of pathos-appealing Argument Schemes seems to be a common rhetorical strategy in political debates. We observe Arguments from Consequences (Positive, Negative) to be the most frequently employed types of arguments in the proposed taxonomy of pathos-appealing Argument Schemes (21% and 18.4%, respectively). Similar findings are reported by Lindahl et al. (2019) in the Swedish political debate. Argument Schemes appealing to strong negative emotions and recognised in the literature as scare tactics, i.e., Danger and Fear Appeal (Walton 2013), are present in 7% of natural language arguments collected in our corpora. Studies show these schemes are also commonly employed in political campaigns (Mařarová 2015), although they do not possess a special place in our data (see Fig. 3b.). We suspect it could be a result of some type of an artifact, as more general categories could be recognised more easily and annotated more frequently. The annotation scheme of pathos-appealing Argument Schemes proposed in the study (see Appendix section 2) could be regarded as reliable at 0.3–0.6 Fleiss' κ level. Nonetheless, the annotation of some of the pathos-appealing Argument Schemes seems to be a more challenging task (Fear Appeal) than annotation of other schemes (Negative Consequences). We observe more general categories (Arguments from Positive/Negative Consequences) to be less challenging to annotate by short-trained raters.

Second, we find the use of emotion-eliciting language in almost 95% of the analysed arguments. On average, we identify 2.75 emotion-eliciting words per argument structure (i.e. premise-conclusion pair). Furthermore, we observe argument premises to be more dense in emotion-eliciting language than argument conclusions. The intensity of emotion-eliciting language is quite low, however, and varies from 0.03 to 0.18 (measured on a normalised scale). We observe joy-eliciting language to be the most intensive and disgust-eliciting language the least intensive. These results might be partially explained by the Pollyanna

effect (positivity bias) observed in affective evaluation of language (Boucher and Osgood 1969; Taboada et al. 2011).

Third, in regard to the audience reactions, we observe emotion-expressing language to be predominantly negative (see Table 7). Conducted correlation analysis reveals that the usage of Arguments from Negative Consequences is associated with the presence of anger-expressing language on social media (see Fig. 5). Appeals to negative emotions were found to be a common strategy in political discussions—Maďarová (2015) finds frequent usage of fear mongering narratives in political debates about gender equality and the human rights of LGBT individuals in Slovakia. Therefore, in our study we extend previous findings on political argumentation by adding analysis of relationship between appeals to emotions by the means of argument structure and reactions to those appeals. Furthermore, we do so also for the usage of emotion-eliciting language and emotion-expressing language on social media. We find several statistically significant associations here. First, we observe that appeals to negative emotions by the means of emotion-eliciting language are correlated with the audience's usage of language expressing negative emotions. Here, we observe a strong relationship between anger-eliciting language and sadness-expressing language (see Table 9). Second, positive (negative) emotion-eliciting language tends to be negatively correlated with negative (positive) emotion-expressing language. However, we also observe a positive association between joy-eliciting language in arguments and disgust-expressing language in the audience response on social media. These findings expose the complex nature of emotional experience, reported also in studies with visual stimuli (Saganowski et al. 2022).

The proposed combination of methods allowed us to overcome certain deficiencies present in the disciplines we borrowed them from. Argumentation theory and modern rhetoric can provide rich theoretical models of *pathos*, where the concept of speaker, audience and the dialogical nature of appeals are clearly described. Psychological method of using laboratory conditions and statistical modelling allows for the generalisation of the emotion types elicited by certain stimuli. Computational linguistics provides models trained on large-scale datasets, which can assess the emotions expressed in a given text span with reasonable accuracy. Yet, it seems like argumentation theory could still benefit from incorporating psychological understanding of emotions, in order to conceive a full concept of *pathos*, suitable for modern rhetoric. Computational linguistics, on the other hand, often misses the theoretical background of interaction structure, rarely distinguishing between appeals to emotion and expressed emotions. Finally, psychological experiments on persuasion could benefit from a more rigid concept of argument as well as using more real-life cases of natural argumentation. The complex issue of *pathos* calls for an interdisciplinary approach, and the results presented in this paper hopefully constitute a step in this direction.

Future studies in a more controlled environment are needed to establish the persuasive role of *pathos* in argumentation. In the current study we were able to test the relation between appeals to emotions by the means of argument structure and emotion-eliciting words on the one side (the speaker), and emotion expression on the other side (the audience) using (semi)-automated methods and large scale samples of natural language data. Perlocutionary effect of appeals to *pathos* in

argumentation could be further investigated in an offline environment in face-to-face interactions and with the use of more traditional forms of research (psychological questionnaires). The possibility of controlling certain variables and manipulating others will allow us to experimentally verify findings presented in the current study. We believe research methods from psychology will allow to discover detailed dependencies between emotion-eliciting and emotion-expressing as well as the usage of emotion-eliciting and occurrence of rhetorical effects in the audience.

6 Conclusions

Aristotelian view on how stirring emotions in the audience can support the rhetorical gain of the speaker is still true in modern-day discourse. In this paper, we propose an updated model of pathos, understood as an interactional persuasive process in which speakers are performing pathos appeals and the audience experiences emotional reactions. We restrict its use to the persuasive context, i.e. the situation in which speakers are aiming at a rhetorical gain, hence the focus on the analysis of those appeals which accompany argumentation. The results point to the importance of pathos analysis in modern discourse: speakers in political debates refer to emotions in most of their arguments, using pathos-appealing argument schemes in 52%, and emotion-eliciting language in 95% of cases. The audience in social media reacts to those appeals using emotion-expressing language, which sometimes is in accordance with speaker's intention (such as reacting with negative sentiment to the use of arguments from Negative Consequences), but sometimes is paradoxical (such as reacting with anger to the appeals to joy). Our results show that pathos is a common strategy in natural language argumentation, however not a straightforward one. We believe that the model of pathos and its operationalisation proposed in this paper paves the way for further analysis of this phenomenon. This study brings empirical evidence to Walton's seminal claim that emotions indeed have a place in argumentation (1992). We follow scholars like Gilbert (2004) who assert the presence of emotional appeals in everyday discourse. While our methodologies are computational, the insights they yield have broader applicability. Our data on the prevalence and types of emotional appeals can guide scholars working with manual discourse analysis, enriching the study of both rational and emotional aspects of argumentation.

While this study offers a comprehensive analysis of the role of pathos in pre-electoral debates, it does have some limitations that opens possibilities of future research. First, our focus on pre-electoral debates inherently limits the scope of the discourse we examine. The importance of emotions in political dialogue raises the question: Are emotions uniquely or disproportionately important in political arenas, or do they play an equally significant role in other spheres of life? Moreover, our data sources (X, prev. Twitter, and Reddit), are not fully representative of the broader population. These platforms attract specific demographics, which might not necessarily capture the wide range of emotions and argument types found in other social groups or platforms. As a result, the findings may not be generalizable to a more diverse audience. Academic study of argumentation has too often concentrated on the language used

by professionals such as politicians, lawyers, and academics. Yet, every day, millions of arguments are formed across languages and cultures. People engage in persuasive discourse in various settings: at the marketplace, in workplaces, in healthcare facilities. The arguments range from mundane choices like buying a kettle to life-altering decisions like getting vaccinated or voting. To fully understand the role of emotional appeals in argumentation, in future studies, we must expand our focus beyond the professional sphere.

Appeals to emotions have accompanied argumentation since the dawn of the rhetoric, and they will continue doing so in the new era of communication in the digital media. Eliciting fear, expressing anger, promising happiness—all these pathotic strategies are contributing to the phenomena observed in social media: hate speech, cyberbullying, the spread of fake news. Argumentation studies can provide rich theoretical framework for analysis of such rhetoric, and with the support of computational methods, will allow for better understanding of pathos in natural language argumentation.

Appendix

Algorithms

Algorithm 1 Emotion-eliciting words

Input: $V \Leftarrow [v_1, \dots, v_i]$ where v_i is a value for the i th word in a text sample assigned to the word in the lexicon
Output: \bar{v}
 1: $\bar{v} \leftarrow \frac{1}{|V|} \sum_{v_i \in V} v_i$

Annotation Scheme

See Fig. 6.

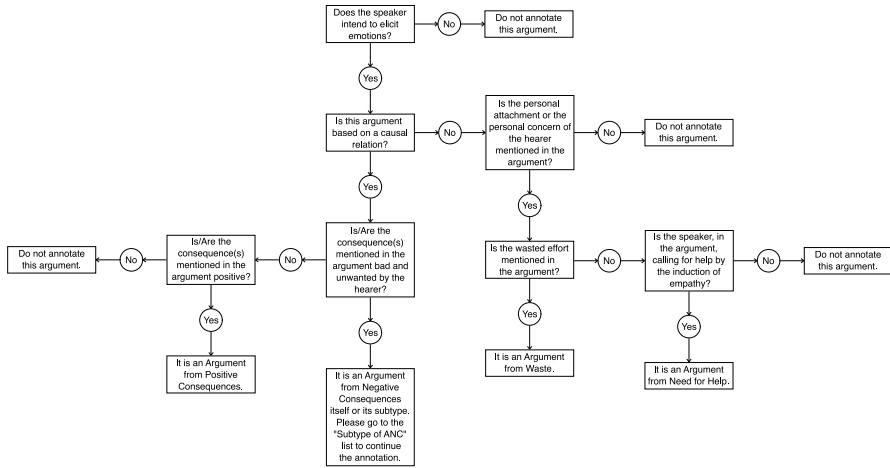


Fig. 6 The proposed annotation scheme of pathotic arguments

Pathotic Argument Schemes Corpora

See Table 10.

Table 10 Corpora re-annotated with pathos-appealing Argument Schemes

Source corpus	Re-annotated corpus	Number of maps	Link
US2016R1tv	US2016R1tv pathotic arguments	40	http://corpora.aifdb.org/R1Pathos
US2016G1tv	US2016G1tv pathotic arguments	26	http://corpora.aifdb.org/G1Pathos
Polish presidential pre-election debates TVP May 2020	Polish presidential pre-election debates TVP May 2020 pathotic arguments	60	http://corpora.aifdb.org/TVPMay2020Pathos
Polish presidential pre-election debates TVP June 2020	Polish presidential pre-election debates TVP June 2020 pathotic arguments	64	http://corpora.aifdb.org/TVPJune2020Pathos

Funding This research was funded in whole by Polish National Science Centre under grant 2020/39/D/HS1/00488.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abercrombie, G., V. Basile, S. Tonelli, V. Rieser, and A. Uma. 2022. Proceedings of the 1st workshop on perspectivist approaches to nlp @LREC2022. In *Proceedings of the 1st workshop on perspectivist approaches to @lrec2022*.
- Alessia, D., F. Ferri, P. Grifoni, and T. Guzzo. 2015. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications* 125 (3): 26–33.
- Alsaedi, A., P. Brooker, F. Grasso, and S. Thomason. 2022. Improving social emotion prediction with reader comments integration. In *Proceedings of the 14th international conference on agents and artificial intelligence*.
- Alsawaidan, N., and M.E.B. Menai. 2020. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems* 62 (8): 2937–2987.
- Aristotle. 2004. *Rhetoric* (W. R. Roberts, Trans.). Dover Publications.
- Artetxe, M., and H. Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics* 7: 597–610.
- Austin, J.L. 1975. *How to do things with words*. Oxford: Oxford University Press.
- Barbieri, F., L.E. Anke, and J. Camacho-Collados. 2021. XLM-T: A multilingual language model toolkit for Twitter. [arXiv:2104.12250](https://arxiv.org/abs/2104.12250).
- Berenguères, J., and D. Castro. 2017. Differences in emoji sentiment perception between readers and writers. In *2017 IEEE international conference on big data (Big Data)*, pp. 4321–4328.
- Boucher, J., and C.E. Osgood. 1969. The Pollyanna hypothesis. *Journal of verbal learning and verbal behavior* 8 (1): 1–8.
- Bourse, S. 2019. Conjuring up terror and tears: the evocative and persuasive power of loaded words in the political discourse on drug reform. *Lexis. Journal in English Lexicology*. <https://doi.org/10.4000/lexis.3182>.
- Braet, A.C. 1992. Ethos, pathos and logos in Aristotle’s Rhetoric: a re-examination. *Argumentation* 6 (3): 307–320.
- Budzynska, K., and C. Reed. 2011a. Speech acts of argumentation: inference anchors and peripheral cues in dialogue. In *Computational models of natural argument: papers from the 2011 AAAI workshop*, pp. 3–10.
- Budzynska, K., and C. Reed. 2011. *Whence inference*. Dundee: University of Dundee Technical Report.
- Buechel, S., and U. Hahn. 2017. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics*, Association for Computational Linguistics, 2: 578–585.
- Cabrio, E., and S. Villata. 2018. The SEEMPAD dataset for emphatic and persuasive argumentation. In E. Cabrio, A. Mazzei, and F. Tamburini (Eds.), *Proceedings of the fifth italian conference on computational linguistics (CLIC-IT 2018)*, Torino, Italy, (Vol. 2253). CEURWS. org. Accessed 10–12 Dec 2018.
- Chatterjee, A., K.N. Narahari, M. Joshi, and P. Agrawal. 2019. SemEval-2019 task 3: EmoContext contextual emotion detection in text. In *Proceedings of the 13th international workshop on semantic evaluation*, pp. 39–48.
- Cigada, S. 2019. Emotions in argumentative narration: the case of the Charlie Hebdo attack. *Informal Logic* 39 (4): 401–431.
- Demszky, D., D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi. 2020. GoEmotions: a dataset of fine-grained emotions. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 4040–4054.
- Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies*. Association for Computational Linguistics, 1: 4171–4186. <https://doi.org/10.18653/v1/N19-1423>

- Diakopoulos, N.A., and D.A. Shamma. 2010. Characterizing debate performance via aggregated Twitter sentiment. In *Proceedings of the Sigchi conference on human factors in computing systems*, pp. 1195–1198.
- Eemeren, F.H.V. 2018. *Argumentation theory: a pragma-dialectical perspective*. Berlin: Springer.
- Ekman, P. 1999. Basic emotions. In *Handbook of cognition and emotion*, ed. T. Dalgleish and M. Power, 45–60. Hoboken: Wiley Online Library.
- Eugenio, B.D., and M. Glass. 2004. The kappa statistic: a second look. *Computational linguistics* 30 (1): 95–101.
- Gilbert, M.A. 2004. Emotion, argumentation and informal logic. *Informal Logic*. <https://doi.org/10.22329/il.v24i3.2147>.
- Gordon, M.L., K. Zhou, K. Patel, T. Hashimoto, and M.S. Bernstein. 2021. The disagreement deconvolution: bringing machine learning performance metrics in line with reality. In *Proceedings of the 2021 chi conference on human factors in computing systems*, pp. 1–14.
- Greco, S., S. Cigada, and C. Jermini-Martinez Soria. 2022. The naming of emotions in dispute mediators' strategic manoeuvring: a case study using a French language corpus. *Text & Talk*. <https://doi.org/10.1515/text-2021-0044>.
- Herman, T., and D. Serafis. 2019. Emotions, argumentation and argumentativity: Insights from an analysis of newspapers headlines in the context of the Greek crisis. *Informal Logic* 39 (4): 373–400.
- Hidey, C., E. Musi, A. Hwang, S. Muresan, and K. McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th workshop on argument mining*, pp. 11–21. Copenhagen, Denmark: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/W17-5102> <https://doi.org/10.18653/v1/W17-5102>
- Hinton, M., and A. Budzysłowska-Daca. 2019. A comparative study of political communication in televised pre-election debates in Poland and the United States of America. *Research in Language* 17 (1): 1–19.
- Janier, M., J. Lawrence, and C. Reed. 2014. OVA+: an argument analysis interface. In *Computational models of argument*, vol. 266, ed. S. Parsons, N. Oren, C. Reed, and F. Cerutti, 463–464. Amsterdam: IOS Press.
- Jose, R., and V.S. Choorailil. 2015. Prediction of election result by enhanced sentiment analysis on Twitter data using word sense disambiguation. In *2015 international conference on control communication & computing India (ICCC)*, pp. 638–641.
- Kissler, J., R. Assadollahi, and C. Herbert. 2006. Emotional and semantic networks in visual word processing: insights from ERP studies. *Progress in brain research* 156: 147–183.
- Kocoić, J., A. Figas, M. Gruza, D. Puchalska, T. Kajdanowicz, and P. Kazienko. 2021. Offensive, aggressive, and hate speech analysis: from data-centric to human-centered approach. *Information Processing & Management* 58 (5): 102643.
- Konat, B., J. Lawrence, J. Park, K. Budzyska, and C. Reed. 2016. A corpus of argument networks: using graph properties to analyse divisive issues. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*.
- Koszowy, M., K. Budzyska, M. Pereira-Fariña, and R. Duthie. 2022. From theory of rhetoric to the practice of language use: the case of appeals to ethos elements. *Argumentation* 36: 123–149.
- Lawrence, J., F. Bex, C. Reed, and M. Snaithe. 2012. AIFdb: infrastructure for the argument web. In *Computational models of argument*, vol. 245, ed. B. Verheij, S. Szeider, and S. Woltran, 515–516. Amsterdam: IOS Press. <https://doi.org/10.3233/978-1-61499-111-3-515>.
- Lindahl, A., L. Borin, and J. Rouces. 2019. Towards assessing argumentation annotation—a first step. In *Proceedings of the 6th workshop on argument mining*, pp. 177–186.
- Liu, H., S. Li, G. Zhou, C.-R. Huang, and P. Li. 2013. Joint modeling of news reader's and comment writer's emotions. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (volume 2: Short papers), pp. 511–515.
- Lukin, S., P. Anand, M. Walker, and S. Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics*, long papers. Association for Computational Linguistics, 1: 742–753.
- Maďarová, Z. 2015. Love and fear argumentative strategies against gender equality in Slovakia. *Anti-Gender movements on the rise* 39: 33–42.
- Manolescu, B.I. 2006. A normative pragmatic perspective on appealing to emotions in argumentation. *Argumentation* 20 (3): 327–343.
- Manolescu, B.I. 2007. Shaming in and into argumentation. *Argumentation* 21 (4): 379–395.

- Micheli, R. 2010. Emotions as objects of argumentative constructions. *Argumentation* 24 (1): 1–17.
- Miłkowski, P., M. Gruza, K. Kanclerz, P. Kazienko, D. Grimling, and J. Kocoń. 2021. Personal bias in prediction of emotions elicited by textual opinions. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: Student research workshop*, pp. 248–259.
- Mohammad, S., F. Bravo-Marquez, M. Salameh, and S. Kiritchenko. 2018. Semeval-2018 task 1: affect in tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pp. 1–17.
- Mohammad, S., and P. Turney. 2010. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pp. 26–34.
- Plantin, C. 2019. Tense arguments: questions, exclamations, emotions. *Informal Logic* 39 (4): 347–371.
- Plutchik, R. 2001. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* 89 (4): 344–350.
- Plutchik, R. 2003. *Emotions and life: perspectives from psychology, biology, and evolution*. Washington: American Psychological Association.
- Russell, J.A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (6): 1161.
- Saganowski, S., Komoszyńska, J., M. Behnke, B. Perz, D. Kunc, B. Klich, and P. Kazienko. 2022. Emognition dataset: emotion recognition with self-reports, facial expressions, and physiology using wearables. *Scientific Data* 9 (1): 1–11.
- Santibáñez, C. 2010. Metaphors and argumentation: the case of Chilean parliamentarian media participation. *Journal of Pragmatics* 42 (4): 973–989.
- Saravia, E., H.-C.T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen. 2018. Carer: contextualized affect representations for emotion recognition. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 3687–3697.
- Searle, J.R. 1979. *Expression and meaning: studies in the theory of speech acts*. Cambridge: Cambridge University Press.
- Stede, M. 2020. Automatic argumentation mining and the role of stance and sentiment. *Journal of Argumentation in Context* 9 (1): 19–41.
- Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics* 37 (2): 267–307.
- Tang, Y.-J., and H.-H. Chen. 2011. Emotion modeling from writer/reader perspectives using a microblog dataset. In *Proceedings of the workshop on sentiment analysis where AI meets psychology (SAAIP 2011)*, pp. 11–19.
- Tang, Y.-J., and H.-H. Chen. 2012. Mining sentiment words from microblogs for predicting writer-reader emotion transition. In *Lrec* pp. 1226–1229.
- Van Haaften, T. 2019. Argumentative strategies and stylistic devices. *Informal Logic* 39 (4): 301–328.
- Villata, S., E. Cabrio, I. Jraidi, S. Benlamine, M. Chaouachi, C. Frasson, and F. Gandon. 2017. Emotions and personality traits in argumentation: an empirical evaluation. *Argument & Computation* 8 (1): 61–87.
- Visser, J., B. Konat, R. Duthie, M. Koszowy, K. Budzynska, and C. Reed. 2020. Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation* 54 (1): 123–154.
- Visser, J., J. Lawrence, C. Reed, J. Wagemans, and D. Walton. 2021. Annotating argument schemes. *Argumentation* 35 (1): 101–139.
- Walton, D. 1992. *The place of emotion in argument*. University Park: The Pennsylvania State University Press.
- Walton, D. 1995. Appeal to pity: a case study of the argumentum ad misericordiam. *Argumentation* 9 (5): 769–784.
- Walton, D. 2007. Evaluating practical reasoning. *Synthese* 157 (2): 197–240.
- Walton, D. 2007. *Media argumentation: dialectic, persuasion and rhetoric*. Cambridge: Cambridge University Press.
- Walton, D. 2010. The structure of argumentation in health product messages. *Argument & Computation* 1 (3): 179–198.
- Walton, D. 2013. *Scare tactics: arguments that appeal to fear and threats*, vol. 3. Berlin: Springer.
- Walton, D., C. Reed, and F. Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511802034>.

- Wang, X., W. Jiang, and Z. Luo. 2016. 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*, pp. 2428–2437.
- Warriner, A.B., and V. Kuperman. 2015. Affective biases in English are bi-dimensional. *Cognition and Emotion* 29 (7): 1147–1167.
- Wierzba, M., M. Riegel, J. Kocoń, P. Milkowski, A. Janz, K. Klessa, et al. 2021. Emotion norms for 6000 Polish word meanings with a direct mapping to the Polish wordnet. *Behavior Research Methods* 54: 2146–2161.
- Wierzba, M., M. Riegel, M. Wypych, K. Jednoróg, P. Turnau, A. Grabowska, and A. Marchewka. 2015. Basic emotions in the Nencki Affective Word List (NAWL BE): new method of classifying emotional stimuli. *PLOS ONE* 10 (7): e0132305.
- Wolf, T., L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, and A.M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pp. 38–45. Online: Association for Computational Linguistics.
- Yang, C., K.H.-Y. Lin, and H.-H. Chen. 2009. Writer meets reader: emotion analysis of social media from both the writer's and reader's perspectives. In *2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology*, 1: 287–290.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.