ARTICLE

LEP Analytics Version 01: User Manual
April 2024: LEPAn v01

Developed on the PolarIs and US2016reddit corpus[†]
Described in the *Routledge* and *COMMA* papers[‡]
For projects: iTRUST

Ewelina Gajewska[*]

Laboratory of The New Ethos,
Faculty of Administration and Social Sciences, Warsaw University of Technology,
Plac Politechniki 1, 00-661 Warsaw, Poland
[*]Corresponding author. Email: ewelina.gajewska.dokt@pw.edu.pl

Abstract

The LEPAn: **L**ogos-**E**thos-**P**athos **An**alytics allows to analyse annotation of selected rhetorical strategies from the Aristotelian triad [1]. We offer both a quantitative and qualitative analysis of appealing to logos, ethos and pathos on social media with the use of statistical and computational methods. Annotation of targets of ethotic appeals allows to further analyse patterns of usage of rhetorical strategies towards individuals and distinguish certain behavioural tendencies. LEPAn analytics are inspired by Argument analytics tools [4]. This user manual describes the first version of the tool (LEPAn v01). It is available at **https://newethos.org/technologies/**.

Keywords: logos – ethos – pathos, trust, communication (mis-)behaviour, argumentation, emotions, public debates, online media, social media

[‡]Gajewska, E., Budzynska, K., Konat, B., Koszowy, M., Kiljan, K., Uberna, M., & Zhang, H. (2024). Ethos and Pathos in Online Group Discussions: Corpora for Polarisation Issues in Social Media., *arXiv preprint, arXiv:2404.04889*, DOI 10.48550/arXiv.2404.04889.

Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (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.

[‡]Katarzyna Budzynska, Marcin Koszowy, Ewelina Gajewska, Maciej Kulik, Maciej Uberna (2024) A Computational Method for Quantitative Analysis of Ethos. In: A Hess and J E Kjeldsen (Eds) *Ethos, Technology, and AI in Contemporary Society: The Character in the Machine*, Routledge.

Gajewska, E., Budzynska, K., & Konat, B. (2024, September). Analytics for Rhetorical Arguments in Discussions on Polarised Issues. In *10th International Conference on Computational Models of Argument-COMMA 2024*.

Contents

1.   Introduction

Concept of analytics technology.   Rhetoric Analytics is a sense-making tool that provides insights into strategic use of language in argumentative discourse. The technology uses data analytics techniques [5, 6] incorporating visualisations in the form of bar charts and radar charts to represent data in a way easy to comprehend that allows us to observe statistical patterns, trends and tendencies. A *pattern* of X is a repeated or regular occurrence of X, i.e., a sequence of data points that follow a recognisable shape or structure. A *trend* of X is a general direction in which X is developing over time or, in our case, over the course of dispute. A *tendency* of X is similar to trend in that both specify a direction, though a tendency accounts for probable dynamics of X. This allows for large-scale discourse analysis, i.e., we are able to make meaningful interpretations of vast amounts of information on how people actually use rhetorical devices in several corpus-analysed discussions.

Implementation of analytics technology.   Streamlit package and Python were employed for the implementation of the web application. Pandas package was utilised for data manipulation and seaborn and matplotlib for visualisation of results. LEPAn allows to combine several corpora into one and analyse it as a single corpora as well as compare patterns of rhetorical behaviour across corpora. The unit of the analysis is either textual or an entity-based. The app is inspired by Argument Analytics tools [4], which we extend by adding ethotic and pathotic arguments as well as entity-based analytics.

2.   LEP Categories

The LEPAn tool makes use of the Aristotelian rhetoric [1] to examine statistical patterns of argumentation in public debates. Three types of rhetorical arguments are distinguished by Aristotle: (i) logotic, which is fact-based, rational argumentation; (ii) ethotic, which is an argument for or against the character (credibility) of the speaker; and (iii) pathotic, which is an emotion-based argumentation that rests on changing the emotional state of the audience.

2.1   Logos

The annotation of logos follows the theoretical framework of Inference Anchoring Theory (IAT) [2]. Two types of relations between propositions distinguished in IAT are analysed in the LEPAn tool: inference (pro-arguments) and conflict (con-arguments) argumentative relations. In an inference relation, one proposition (a premise of an argument) supports another proposition (a conclusion of an argument). This pro-argumentation is illustrated by Ex. (1-a), where senator Paul justifies his visits to Chicago, Detroit, Ferguson and Baltimore. When two propositions cannot be true at the same time, there is a conflict relation between these propositions. Ex. (1-b) illustrates logotic con-argumentation: user-2 disagrees with user-1 that Covid-19 vaccines are great. Annotation of both types of logos relations (inference and conflict) are presented in Figure 1. OVA tool[1] is employed for the annotation of logos.

---

1. https://ova.arg.tech/

(1)   a.   Paul: *I want our party to be bigger, better and bolder.* Paul: *I've also gone to Chicago, I've gone to Detroit, I've been to Ferguson, I've been to Baltimore.* [US2016reddit, L+]
       b.   USER–1: *COVID 19 vaccine is great.* USER–2: *Data shows they are not so great.* [PolarIs1, L–]
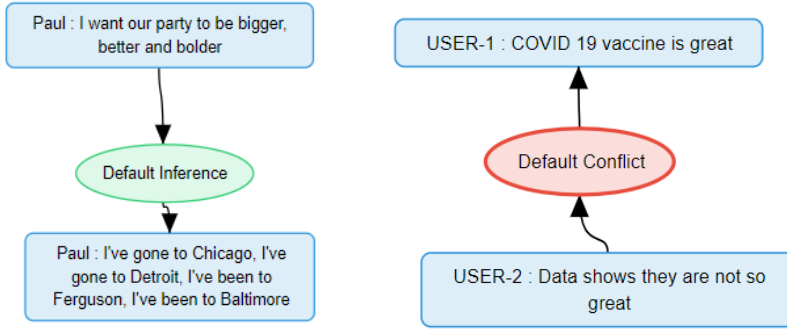

Figure 1. Example of logos annotation: inference and conflict relations.

2.2   Ethos

We follow a redefinition of the traditional conceptualisation of ethotic arguments, and regard ethos as a speaker's property, which can be attacked or supported by other speakers [3]. Thus, we treat favourable (positive) references to a speaker (a person, a group of persons or an organisation) as ethotic supports and unfavourable (negative) references as ethotic attacks. Ex. (2-a)-(2-b) illustrate the former and Ex. (2-c)-(2-d) the latter type of ethotic arguments. An excel sheet with the annotation of ethotic arguments is presented in Figure 2.

(2)   a.   USER–3: *Yes you are very correct.* [US2016reddit, E+]
       b.   USER–4: *Your vigilance is commendable (and should be the norm).* [PolarIs1, E+]
       c.   USER–5: *You're misinterpreting data to promote your own bias.* [PolarIs1, E–]
       d.   USER–6: *Reading / watching CNN is like eating junk food out of the trash.* [PolarIs1, E–]

2.3   Pathos

Pathos is the emotional state of the audience that speakers attempt to influence through their argumentation **r7**. Pathotic argument can appeal to either positive or negative emotions – examples of such an argumentation is presented in Ex. (3-a)-(3-b) regarding the former and in Ex. (3-c)-(3-d) regarding the latter types of pathotic arguments. Similarly to ethos (Figure 3), pathos annotation is conducted in excel sheets.

(3)   a.   USER–7: *Without progress we don't improve as a country or as a people.* [US2016reddit, P+]

	G	H	I	J	K	L
1	sentence	No_Ethos	Contains_ethos	Support	Attack	Target
324	[deleted] : they've had 8 years to find one guy to do that	1	0			
325	EarthExile : Are you still laughing as three more people also s	1	0			
326	socialistbob : He will flame out soon	0	1	0	1	Trump
327	socialistbob : This is the time when Herman Cain and Michell	1	0			
328	Giant_Asian_Slackoff : many Americans wonder the same thi	1	0			
329	Giant_Asian_Slackoff : Please don't judge us	1	0			
330	Giant_Asian_Slackoff : I swear we're not all crazy	1	0			
331	WALLACE : Mr. Trump, I'm giving you 30 seconds to answer n	1	0			
332	parduscat : WALLACE : Mr. Trump, I'm giving you 30 seconds	1	0			
333	parduscat : Fox News is out for Trump's blood	0	1	0	1	Fox News
334	dirtyfries : Yep	1	0			
335	dirtyfries : I hate the guy	0	1	0	1	Trump
336	dirtyfries : it's clear this whole thing is about picking winners	1	0			
337	dirtyfries : They want him gone	0	1	0	1	Fox News
338	lolamp33 : no one can accuse fox news of playing soft ball	0	1	1	0	Fox News
339	lolamp33 : I dont think hillary is going to get the same treatm	0	1	0	1	Fox News
340	lolamp33 : shes the queen	0	1	1	0	Clinton
341	liamliam1234liam : I really hope they pull no punches	1	0			

Figure 2. Example of ethos annotation in an excel sheet.

b.   USER–8: *It's pulled hundreds of millions of people out poverty around the world.* [US2016reddit, P+]

c.   USER–9: *But also – ignoring that a huge proportion of the population have comorbidities (you might have one yourself), yet saying 'fuck them' is pretty disgusting.* [PolarIs1, P–]

d.   USER–10: *Playing the "people I know died in 9/11" card was pretty fucking low.* [US2016reddit, P–]

	B	C	D	E	F
1	sentence	No_pathos	Contains_pathos	positive_valence	negative_valence
324	they've had 8 years to find one guy to do that	0	1	0	1
325	Are you still laughing as three more people also support the Wall	1	0	0	0
326	He will flame out soon	1	0	0	0
327	This is the time when Herman Cain and Michelle Bachmann were leading last time around	1	0	0	0
328	many Americans wonder the same thing	1	0	0	0
329	Please don't judge us	1	0	0	0
330	I swear we're not all crazy	1	0	0	0
331	Mr. Trump, I'm giving you 30 seconds to answer my question	1	0	0	0
332	WALLACE : Mr. Trump, I'm giving you 30 seconds to answer my question	1	0	0	0
333	Fox News is out for Trump's blood	0	1	0	1
334	Yep	1	0	0	0
335	I hate the guy	0	1	0	1
336	it's clear this whole thing is about picking winners and losers	1	0	0	0
337	They want him gone	1	0	0	0
338	no one can accuse fox news of playing soft ball	0	1	0	1
339	I dont think hillary is going to get the same treatment in the dem debates	1	0	0	0
340	shes the queen	0	1	1	0

Figure 3. Example of pathos annotation in an excel sheet.

3.   Interface Structure

3.1   Sidebar/Navigation

On the left hand–side of the interface, there is a sidebar for filtering data and specifying conditions of analytics one wants to display (marked by 1 in Figure 4).

3.2   Main Panel

On the right hand–side/centre of the interface, description and results of the analysis are presented (the rectangle marked by 2 in Figure 4).
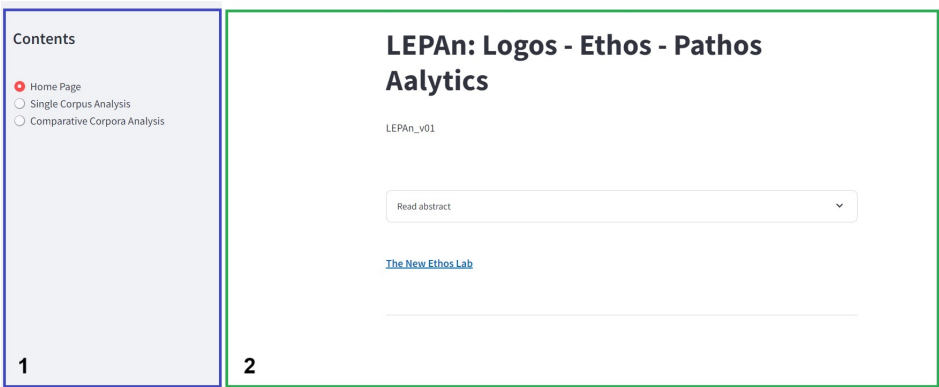

Figure 4. The structure of the interface.

3.3   Map of Interface

Figure 5. Interface map.

- 1: Type of the conducted corpus analysis – Single versus Comparative Corpora Analysis;

- 2: Corpora available for the analysis – the end user can choose one (in single corpora analysis) or more than one (single and comparative corpora analysis) corpus for the analysis. In the 'Single Corpus Analysis' scenario these corpora will be merged and treated as one corpus comprising of many subcorpora; in the 'Comparative Corpora Analysis' condition a separate analysis will be conducted for each corpus and results will be displayed next to each other in the main interface;
- 3: Unit for the analysis: text-based (a piece of text is the basic unit for analysis) versus entity-based (a user, a target, or a speaker and text content related to them is the basic unit of analysis);
- 4: Statistical modules that will be displayed on the main interface;
- 5: The unit of the calculated statistics (in percentages or numbers);
- 6: The type of categories/labels taken for the calculation of statistics;
- 7: Tabs: results of the calculated statistic are presented in different forms and displayed in separate tabs (for example, as a bar chart, pie chart and table);
- 8: Visualisation of the calculated statistics.

4. Interface Use

4.1 Home page

It introduces the end user to the analysis available in the LEPAn tool. Abstract of the paper that describes and discusses the idea and results of LEP analytics is displayed. A summary of the categories employed in the LEP analytics (logos, ethos, pathos) is presented.

4.2 Corpus Analysis

4.2.1 Single corpus analysis

In this type of analysis the user chooses corpora from the available set of corpora which are combined into a single corpus and treated as such by all statistical modules. That is, if a user chooses PolarIs1 and US2016reddit corpora, they will be merged into a single corpus named henceforth PolarIs1 & US2016reddit in the analytics.

4.2.2 Comparative corpora analysis

The second type of analysis allows to compare tendencies in rhetorical behaviour between corpora – the user chooses at least two corpora; statistics for each of them are calculated separately. That is, if a user chooses PolarIs1 and US2016reddit corpora, for example in the distribution module, density of use of logos, ethos and pathos will be calculated separately for PolarIs1 and for US2016reddit corpora.

4.3 Choose corpora

In the side-bar on the left side of the interface, the user chooses corpora for analysis. The choice is conducted by ticking the box next to the name of the corpus. For example, in Figure 5 two corpora are available and the two corpora are chosen for a statistical analysis.

In the first version of the LEPAn tool, two corpora are available: Reddit discussions on Covid–19 vaccines (PolarIs1) and Reddit reactions to live TV debates from the US 2016 presidential elections (US2016reddit). Summary of corpora size is presented in Table 1. Summary of categories used in LEP analytics is presented in Table 2.

Table 1. Datasets used in our technology of LEP Analytics.

Corpus	Words	ADU	Posts	Speakers
PolarIs1	30,014	2,706	963	465
US2016reddit	30,099	3,827	1,317	1,317
Total	60,113	6,533	2,280	1,782

Table 2. Annotation of logos, ethos and pathos used in LEP Analytics.

Annotation	Logos			Ethos			Pathos		
Corpus	**L-**	**L+**	**IAA**	**E-**	**E+**	**IAA**	**P-**	**P+**	**IAA**
PolarIs1	630	1233	0.618	440	59	0.752	653	152	0.417
US2016reddit	581	1144	0.817	847	492	0.793	1294	190	0.573
Total/Average	1211	2377	0.718	1287	551	0.773	1947	342	0.495

4.4   Analysis units

LEPAn tool offers two units of analysis: **text**-based and **entity**-based. In the former, a piece of text (a sentence, a tweet, a post) is the basic unit of analysis – for example, in the Distribution module, number of sentences labelled as (either positive or negative) logos, ethos and pathos are counted. We offer two statistical modules for text-based analytics: distribution and wordcloud. In the entity-based analysis, a user, a target, or a speaker is the unit of analysis. For example, in the Heroes & villains Score–1 module, information about ethotic profiles of entities (proportion between ethotic attacks and supports) mentioned in a discussion is displayed. Entity-based analytics investigate the image of speakers or targets of emerged from the public discussion. We call this target-based analytic module 'heroes & villains': it allows to identify entities mentioned in a given discussion that the public views as heroes (i.e., the public mostly supports their ethos) and villains (i.e., the public mostly attacks their ethos). Such a public image of entities is obtained with the annotation of ethotic expressions in text, thus we call it *the ethotic profile* of an entity. These ethotic profiles are expressed numerically in the LEPAn tool with our concept of villain scores. Heroes & villains module present results of the analysis of villain scores in two forms: plots and tables, which are presented in a separate tabs on the interface. We offer four statistical modules for text–based analytics: heroes & villains frequency, score, profiles and wordcloud.

4.5   Statistical Modules

4.5.1 Distribution

A bar chart with percentage distribution of logos, ethos and pathos categories is displayed. Three configurations of LEP categories are available: 3–LEP categories, 6–LEP categories, 4–E categories. 3–LEP categories involve calculation of density of logos, ethos and pathos arguments in the chosen corpora. 6–LEP categories calculates proportions between pro– (supporting) arguments and con– (attacking) arguments for LEP categories. Finally, 4–E categories report proportions between ethotic supports and attacks on direct versus 3rd party entities. Direct ethos refers to entities that can directly participate in a discussion on an online platform, i.e., to social media users. Third party ethos in turn covers cases where actors from the outside world are just mentioned in a discussion and do not actively participate in the discussion. Figure 6 presents a bar–chart with logos, ethos and pathos density in corpora. LEP distribution is visualised in separate tabs also as a pie–chart and reported in a table. This statistical module allows to compare the usage of these categories across rhetorical strategies of logos, ethos and pathos, and across corpora. In addition, one can compare how the type of targets of ethotic arguments influences distribution of support versus attack categories.
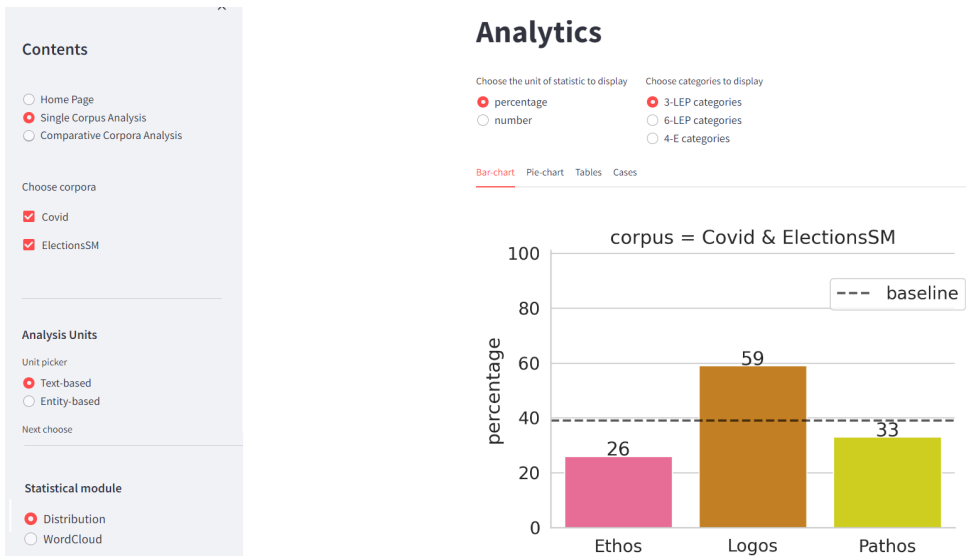

Figure 6. Text-based Analytics: Distribution module.

4.5.2 WordCloud

In this module one can display words from the chosen rhetorical strategy and its category that are above a specified precision value. Precision here refers to the percentage of times we find a certain word occurring in text belonging to a given category given all occurrences of this word in the chosen corpora. These words are then visualised as a map of words (in the 'Plots' tab, and listed in a table (the 'Table' tab). A table comprises information about the number of occurrences of a particular term in utterances classified as support (positive),

attack (negative) and neutral as well as an overall number of occurrences of this term in the corpora and its precision score. Lastly, cases from the corpora, where these words are present, are displayed in a dataframe in the 'Cases' tab.
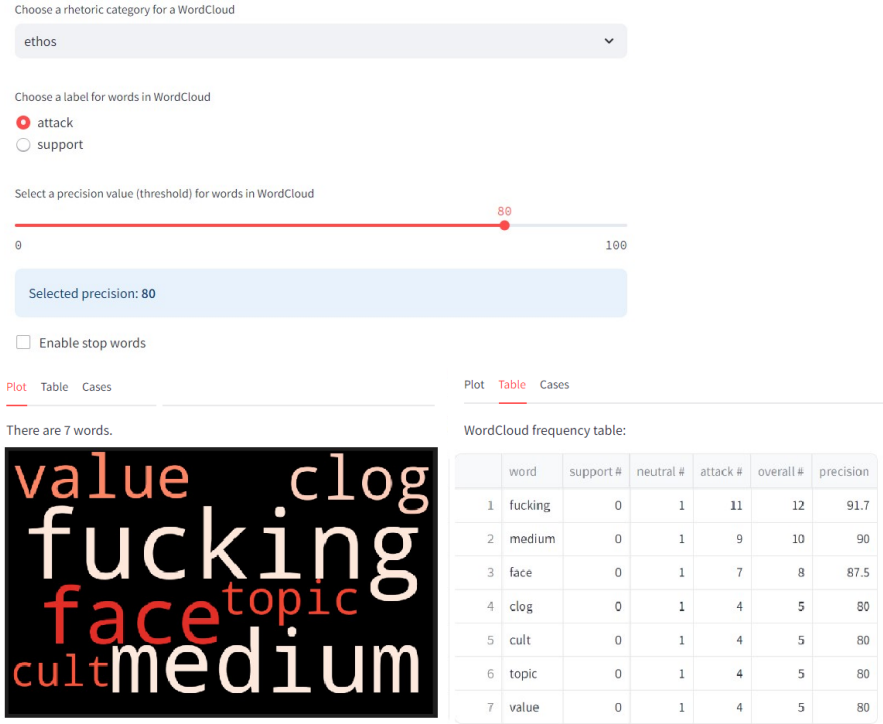

Figure 7. Text-based Analytics: WordCloud module.

4.5.3   Heroes & Villains

Categorisation of entities into heroes and villains is based on the number of supports and attacks towards a particular individual: if the number of attacks is higher than supports of entity e, then entity e is labelled as villain, and hero otherwise. It allows us to position a list of target entities on a spectrum from public heroes (good guys) to public villains (bad guys). The statistic allows to identify most negative and positive entities in a given discussion as well as determine differences in ethotic profiles between entities. Figure 8 shows that both scores classify the same entities as heroes and villains. Both scores identify also Republicans and Fox News as the most negative villains in the US2016reddit corpus. What is different is the relative differences between entities and their positions on a scale. Ethotic profiles of entities are numerically expressed on a scale from –1 to 1, with 0 as the middle point that classifies entities into villains (score < 0) and heroes (score > 0.5). The distinction between heroes and villains is more clearly visible in this case, especially in a graphic presentation of the results in a form of bar–charts as in Figure 8.
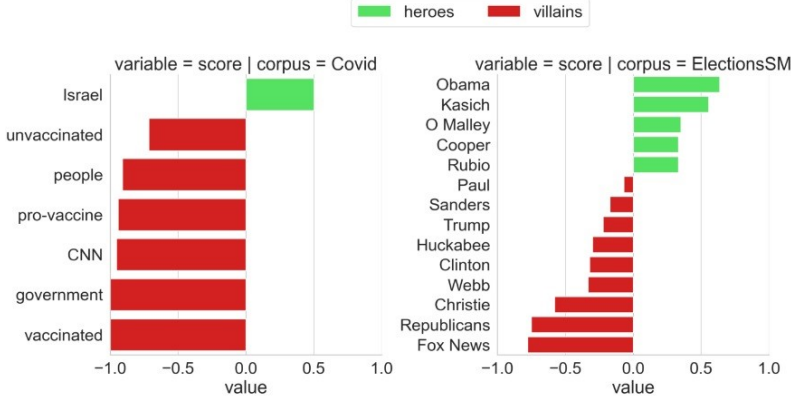
Figure 8. Entity-based Analytics: Heroes & villains.

5.   Formulas

5.1   Distribution

3–LEP categories.   It involves calculation of density of logos, ethos and pathos arguments in the chosen corpora. Thus, annotation if binarised: pro- and con- arguments are aggregated into one label (into contains logos, contains ethos or contains pathos). Then, sentences that do not contain any arguments are labelled as no logos, no ethos or no pathos. Percentage or number of sentences marked with positive and negative categories is calculated (contains logos vs. no logos, contains ethos vs. no ethos and contains pathos vs. pathos). Percentage density is calculated according to the formula 1. Density expressed as a number is just a number of pro and con arguments in a corpora.

$$density = \frac{N_{pro} + N_{con}}{N} \tag{1}$$

where N_{pro} is the number of sentences annonated as (either logos, ethos or pathos) pro–arguments; N_{con} is the number of sentences annonated as (either logos, ethos or pathos) con–arguments; and N is the number of sentences in the corpora.

Baseline is calculated according to the formula 2 as the mean density of LEP categories in the corpora by averaging the densities of logos, ethos and pathos arguments:

$$baseline = \frac{L_{density} + E_{density} + P_{density}}{3} \tag{2}$$

Deviation is calculated as the difference between the density and the baseline:

$$deviation = density - baseline \tag{3}$$

Density and deviation is calculated separately for each category of logos, ethos and pathos; baseline is calculated once for all rhetorical strategies.

6–LEP categories.   6–LEP categories calculates proportions between pro– (supporting) arguments and con– (attacking) arguments for logos, ethos and pathos. Density expressed as a percentage is calculated then according to equation 4:

$$densityPro = \frac{N_{pro}}{N_{pro} + N_{con}}; densityCon = \frac{N_{con}}{N_{pro} + N_{con}} \tag{4}$$

where N_{pro} is the number of sentences annonated as (logos, ethos or pathos) pro–arguments and N_{con} is the number of sentences annonated as (logos, ethos or pathos) con–arguments. Then two baselines are calculated, separately for pro– (supporting) and con– (attacking) arguments with a formula 5:

$$baselinePro = \frac{L_{densityPro} + E_{densityPro} + P_{densityPro}}{3} \tag{5}$$

Baseline for con–arguments is calculated in the same manner – by adding the number of logos, ethos and pathos con–arguments and dividing by 3. Formula for calculating deviation for 6–LEP categories is the same as in case of 3–LEP categories, i.e., it is a subtraction of the baseline from the density for each pro– and con– category. Similarly, density and deviation are calculated separately for each of the 6 categories, baseline is calculated twice: one for pro– categories and one for con– categories.

4–E categories.   4–E categories deals with densities of ethotic pro– and con–arguments that either refer to the ethos of users that take active part in the public discussion (direct) or 3rd party entities that are only mentioned in the discussion but do not actively participate in it (3rd party). Thus, densities are calculated according to equation 4. Baseline is the average density of ethos arguments:

$$baseline = \frac{E_{densityPro} + E_{densityCon}}{2} \tag{6}$$

Deviations are calculated in the same manner as in case of 3– and 6–LEP categories (eq. 3).

5.2 WordCloud

$$precisionWpro = \frac{N_{Wpro}}{N_W} \times 100; precisionWcon = \frac{N_{Wcon}}{N_W} \times 100 \tag{7}$$

where N_{Wpro} is the number of occurrences of word W in pro–arguments, where N_{Wcon} is the number of occurrences of word W in con–arguments, N_W is the total number of occurrences of word W in the corpus.

5.3 Heroes & Villains

Score for an entity e ($score_e$) is calculated according to formula 8 (score–2) based on the number of ethotic attacks on entity e (N_{ae}) and number of ethotic supports of entity e (N_{se}).

$$score_e = \frac{N_{ae} - N_{se}}{N_{ae} + N_{se}} \tag{8}$$

Acknowledgements

References

[1] Aristotle. *Rhetoric*. Dover Publications, 2004.

[2] Budzynska, K., & Reed, C. (2011). Whence inference. *University of Dundee Technical Report*.

[3] Duthie, R., & Budzynska, K. (2018, July). A Deep Modular RNN Approach for Ethos Mining. In *IJCAI* (pp. 4041-4047).

[4] Lawrence, J., Duthie, R., Budzynska, K., & Reed, C. (2016). Argument analytics. In *Computational Models of Argument* (pp. 371-378). IOS Press.

[5] Konat, B., and Gajewska, E., & Rossa, W. (20xx). Pathos in natural language argumentation: emotional appeals and reactions. *Argumentation*. In press.

[6] Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT press.

[7] Walker, R. (2015). *From big data to big profits: Success with data and analytics*. Oxford University Press.