

# Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum

**Christopher Hidey**  
Computer Science Department  
Columbia University  
chidey@cs.columbia.edu

**Elena Musi**  
Data Science Institute  
Columbia University  
em3202@columbia.edu

**Alyssa Hwang**  
Computer Science Department  
Columbia University  
a.hwang@columbia.edu

**Smaranda Muresan**  
Data Science Institute  
Columbia University  
smara@columbia.edu

**Kathleen McKeown**  
Computer Science Department  
Columbia University  
kathy@cs.columbia.edu

## Abstract

Argumentative text has been analyzed both theoretically and computationally in terms of argumentative structure that consists of argument components (e.g., claims, premises) and their argumentative relations (e.g., support, attack). Less emphasis has been placed on analyzing the semantic types of argument components. We propose a two-tiered annotation scheme to label claims and premises and their semantic types in an online persuasive forum, *Change My View*, with the long-term goal of understanding what makes a message persuasive. Premises are annotated with the three types of persuasive modes: *ethos*, *logos*, *pathos*, while claims are labeled as *interpretation*, *evaluation*, *agreement*, or *disagreement*, the latter two designed to account for the dialogical nature of our corpus.

We aim to answer three questions: 1) can humans reliably annotate the semantic types of argument components? 2) are types of premises/claims positioned in recurrent orders? and 3) are certain types of claims and/or premises more likely to appear in persuasive messages than in non-persuasive messages?

## 1 Introduction

Argumentation is a type of discourse where speakers try to persuade their audience about the reasonableness of a claim by displaying supportive arguments. As underlined in Rhetorics and Argumentation Theory (Perelman and Olbrechts-Tyteca, 1973; van Eemeren and Eemeren, 2009),

the persuasiveness of a message lies at the interface between discourse form (i.e., use of hedges, connectives, rhetorical questions) and conceptual form such as the artful use of *ethos* (credibility and trustworthiness of the speaker), *pathos* (appeal to audience feelings), and *logos* (appeal to the rationality of the audience through logical reasoning). Recent work in argumentation mining and detection of persuasion has so far mainly explored the persuasive role played by features related to discourse form (Stab and Gurevych, 2014a; Peldszus and Stede, 2016; Habernal and Gurevych, 2016; Tan et al., 2016; Ghosh et al., 2016). However, due to the lack of suitable training data, the detection of conceptual features is still nascent.

On these grounds, we propose and validate a systematic procedure to identify conceptual aspects of persuasion, presenting a two-stage annotation process on a sample of 78 threads from the sub-reddit *Change My View* (Section 3). *Change My View* constitutes a suitable environment for the study of persuasive argumentation: users award a *Delta point* to the users that managed to changed their views, thus providing a naturally labeled dataset for persuasive arguments. In the first stage, expert annotators are asked to identify claims and premises among the propositions forming the post. In the second stage, using crowdsourcing (Amazon Mechanical Turk) claims and premises are annotated with their semantic types. For premises, the semantic types are based on the Aristotelian modes of persuasion *logos*, *pathos* and *ethos*, or a combination of them. For claims, we have considered two proposition types among those in Freeman’s taxonomy (Freeman, 2000) that can work as claims since their truth is assailable, namely *interpretations* and *evaluations* (*rational/emotional*).

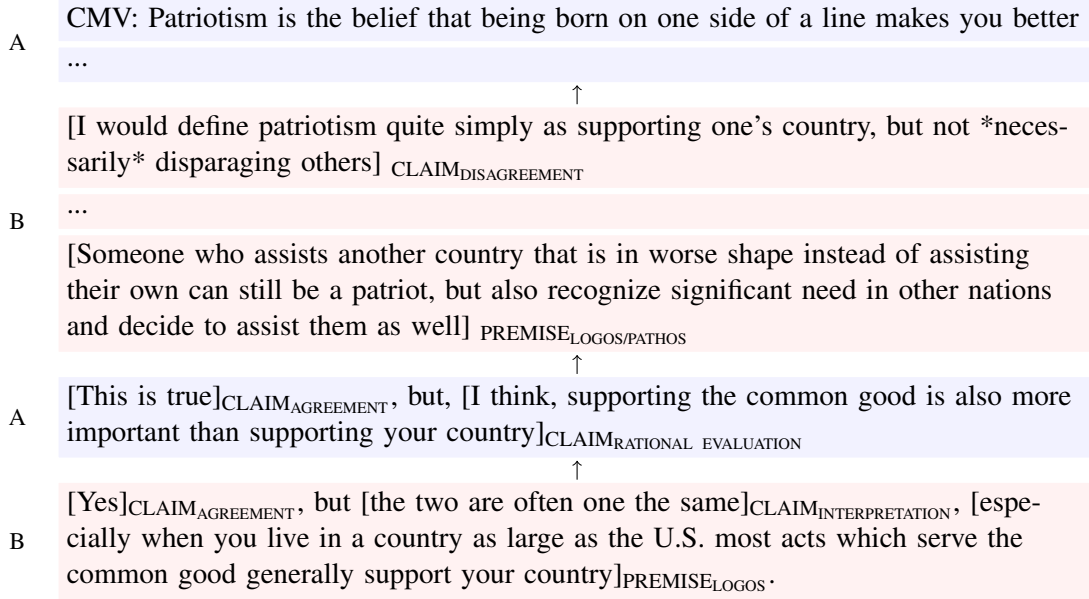


Figure 1: Annotation Example

We have furthermore distinguished propositions expressing *agreement* and *disagreement* because they present an anaphoric function inherent to the dialogic nature of the corpus. An example is given in Figure 1.<sup>1</sup>

We aim to answer three questions: 1) can humans reliably annotate claim and premises and their semantic types? (Section 4) 2) are types of premises/claims positioned in recurrent orders? and 3) are certain types of claims and/or premises more likely to appear in persuasive messages than in non-persuasive messages? (Section 5.2). Our findings show that claims, premises and premise types can be annotated with moderate agreement (Krippendorff's  $\alpha > 0.63$ ), while claim types are more difficult for annotators to reliably label ( $\alpha = 0.46$ ) (Section 4). To answer the second question, we perform an analysis of the correlations between types of argumentative components (premises and claims), as well as their position in the post and discuss our findings in Section 5.1. Our results for the third question show that there are several significant differences between persuasive and non-persuasive comments as to the types of claims and premises (Section 5.2). We present our future work in Section 6. The annotated dataset is available on GitHub to the research community<sup>2</sup>.

<sup>1</sup>Note that premises are labeled at proposition level and not clause level.

<sup>2</sup><https://github.com/chridey/change-my-view-modes>

## 2 Related Work

There are three areas relevant to the work presented in this paper, which we address in turn.

**Persuasion detection and prediction.** Recent studies in argument mining and computational social science have focused on persuasion detection and prediction. A bulk of them have focused on the identification of structural and lexical features that happen to be associated with persuasive arguments. Ghosh et al. (2016) have shown that the number of supported/unsupported claims and the structure of arguments directly affect persuasion. Habernal and Gurevych (2016) have experimented with SVM and bidirectional LSTM to predict arguments scored by annotators as convincing mainly using lexical linguistic features (e.g., modal verbs, verb tenses, sentiment scores). Taking advantage of the *Change My View* dataset, (Tan et al., 2016), have investigated whether lexical features and interaction patterns affect persuasion, finding that lexical diversity plays a major role. In a similar vein, other studies have ranked arguments according to their karma scores (Wei et al., 2016), showing that aspects of argumentative language and social interaction are persuasive features. In this paper, we focus on the conceptual aspects of a persuasive message by analyzing the semantic types of claims and premises. A closely related area of research is the detection of situational influencers — participants in a discussion

who have credibility in the group, persist in attempting to convince others, and introduce ideas that others pick up on or support (Rosenthal and Mckeown, 2017; Biran et al., 2012). In particular, Rosenthal and Mckeown (2017) draw their approach from Cialdini’s (Cialdini, 2005) idea of “weapons of influence,” which include reciprocation (sentiment and agreement components), commitment (claims and agreement), social proof (dialog patterns), liking (sentiment and credibility), authority (credibility), and scarcity (author traits). Our approach zooms into the detection of commitment analyzing not only the presence of claims/arguments, but also their conceptual type. We, moreover, treat credibility as an argument type.

**Modes of persuasion: logos, pathos, ethos.** At the conceptual level, the distinction between different modes of persuasion dates back to Aristotle’s Rhetorics. Aristotle considered that a good argument consists of the contextually appropriate combination of *pathos*, *ethos*, and *logos*. Duthie et al. (2016) have developed a methodology to retrieve *ethos* in political debates. Higgins and Walker (2012) traced back *ethos*, *pathos* and *logos* as strategies of persuasion in social and environmental reports. Their definition of *logos* applies both to premises and claims, while we consider *logos* as referred to arguments only. Habernal and Gurevych (2017) have also included *logos* and *pathos*, but not *ethos*, among the labels for an argumentatively annotated corpus of 990 user generated comments. They obtained moderate agreement for the annotation of *logos*, while low agreement for *pathos*. Our study shows moderate agreement on all types of persuasion modes. On the computational side, the Internet Argument Corpus (IAC) (Walker et al., 2012) — data from the online discussion sites *4forums.com* and *CreateDebate* — includes the distinction between fact and emotion based arguments. Das et al. (2016) looked at the diffusion of information through social media and how author intent affects message propagation. They found that persuasive messages were more likely to be received positively if the emotional or logical components of a message were selected according to the given topic. Lukin et al. (2017) examined how personality traits and emotional or logical arguments affect persuasiveness.

**Semantics of argument components.** Recently, new interest has arisen in analyzing the semantics of argument components. Becker et al. (2016) have investigated correlations between situation entity types and claims/premises. Park et al. (2015) have proposed a classification of claims in relation to the subjectivity/objectivity of the premises in their support. On a different note, a scalable and empirically validated annotation scheme has been proposed for the analysis of illocutionary structures in argumentative dialogues drawing from *Inference Anchoring Theory* (Budzynska et al., 2014; Budzynska and Reed, 2011), relying on different types of *pragmatic* information. However, distinct taxonomies to account for semantic differences characterizing claims vs. premises and their degrees of persuasiveness has so far not been investigated.

Our study contributes to previous work in proposing a novel and reliable annotation scheme, which combines semantic types for both claims and premises at the propositional level, allowing to observe relevant combinations in persuasive messages.

### 3 Annotation Process

#### 3.1 Source data

Change My View is a discussion forum on the site reddit.com. The initiator of the discussion will create a title for their post (which contains the major claim of the argument) and then describe the reasons for their belief. Other posters will respond and attempt to change the original poster’s view. If they are successful, the original poster will indicate that their view was changed by providing a  $\Delta$  point. We use the same dataset from the Change My View forum created in previous work (Tan et al., 2016). We extract dialogs from the full dataset where only the original poster and one responder interacted. If the dialogue ends with the original poster providing a  $\Delta$ , the thread is labeled as positive; if it ends prematurely without a  $\Delta$ , it is labeled negative. We select 39 positive and 39 negative threads to be annotated.

#### 3.2 Annotation of argumentative components

In the first stage of the annotation process, the goal is to label claims and premises at the proposition level. We recruited 8 students with a background either in Linguistics or in Natural Language Processing to be annotators. Students were asked to

read the guidelines and were given an example with gold labels (see Figure 1). During a one-hour long training session they were asked to annotate a pilot example and comparison between their preliminary annotations and the gold labels was discussed. Each student annotated from a minimum of 5 to a maximum of 22 threads depending on their availability.

The guidelines provide an intuitive definition of claims/premises paired with examples. While the definitions are similar to those provided in previous annotation projects (Stab and Gurevych, 2014b), we took as annotation unit the proposition instead of the clause, given that premises are frequently propositions that conflate multiple clauses (see Figure 1).

- **claim:** proposition that expresses the speaker’s stance on a certain matter. They can express predictions (‘I think that the left wing will win the election’), interpretations (‘John probably went home’), evaluations (‘Your choice is a bad one’) as well as agreement/disagreement with other peoples claims (‘I agree’/‘I think you are totally wrong’). Complex sentences where speakers at first agree and then disagree with other speakers’ opinion (*concessions*) constitute separate claims (‘I agree with you that the environmental consequences are bad, but I still think that freedom is more important.’).
- **premise:** proposition that expresses a justification provided by the speaker in support of a claim to persuade the audience of the validity of the claim. Like claims, they can express opinions but their function is not that of introducing a new stance, but that of supporting one expressed by another proposition (‘John probably went home. I don’t see his coat anywhere’; ‘Look at the polls; I think that the right wing will win the election’).

Both claims and premises can be expressed by rhetorical questions, questions that are not meant to require an answer — which is obvious — but to implicitly convey an assertive speech act. Their argumentative role, thus, has to be decided in context: in the sentence ‘We should fight for our privacy on the Web. Dont you love that Google knows your favorite brand of shoes?’, the rhetorical question functions as an argument in support of the recommendation to fight for privacy.

Completely untagged sections mostly contain greetings, farewells, or otherwise irrelevant text. Thus, occasionally entire paragraphs are left unmarked. Furthermore, we left the title unannotated, assuming that it works as the original poster’s major claim, while we are interested in the comments that could persuade the original poster to change his view. When the original poster’s text starts with an argument, it is by default to be considered in support of the title.

### 3.3 Annotation of types of premises and claims

The second stage aims to label the semantic type of claims and premises using crowdsourcing. We used *Amazon Mechanical Turk* (AMT) as our crowdsourcing platform. Using the previous annotations of claim/premises, Turkers were asked to identify the semantic type of premises and claims. The novelty of this study relies in the proposal of a fine-grained, non context-dependent annotation of semantic types of premises and of claims. On the other hand, existing semantic classifications focus either on premises or on claims (section 2). Current Studies have by far tackled types of premises and claims combinations specific to a restricted set of argument schemes (Atkinson and Bench-Capon, 2016; Lawrence and Reed, 2016) mainly for classification purposes.

For each claim, we showed the workers the entire sentence containing the claim. For each premise, we showed the Turkers the entire sentence containing the premise and the sentence containing the claim. Each HIT consisted of 1 premise or 1 claim classification task and the Turkers were paid 5 cents for each HIT.

For claims, the Turkers were asked to choose among four different choices. The distinction between interpretations and evaluations recalls Freeman’s (Freeman, 2000) classification of contingent statements. We have decided to treat agreements/disagreements as distinct types of claims since, depending on the semantics of the embedded proposition, they can express sharedness (or not) of interpretations as well as evaluations. The provided definitions are:

- **interpretation:** expresses predictions or explanations of states of affairs (‘I think he will win the election.’ or ‘He probably went home.’)
- **evaluation:** the claim expresses a more or

less positive or negative judgement. Drawing from the distinction made in sentiment analysis and opinion mining, (Liu, 2012) evaluations are sub-classified as:

- **evaluation-rational**: expresses an opinion based on rational reasoning, non-subjective evidence or credible sources (“His political program is very solid.” or “He is a very smart student.”)
- **evaluation-emotional**: expresses an opinion based on emotional reasons and/or subjective beliefs (“Going to the gym is an unpleasant activity.” or “I do not like doing yoga.”)
- **agreement or disagreement**: expresses that the speaker shares/does not share to a certain degree the beliefs held by another speaker, i.e. “I agree that going to the gym is boring” or “you are right” or “I do not think that he went home.” or “You are not logically wrong.” or “I do not like your ideas.” or “It may be true.”

For premises, the Turkers were provided with the following labels:

- **logos**: appeals to the use of reason, such as providing relevant examples and other kinds of factual evidence (“Eating healthy makes you live longer. The oldest man in the US followed a strictly fat-free diet.” or “He will probably win the election. He is the favorite according to the polls.”)
- **pathos**: aims at putting the audience in a certain frame of mind, appealing to emotions, or more generally touching upon topics in which the audience can somehow identify (“Doctors should stop prescribing antibiotics at a large scale. The spread of antibiotics will be a threat for the next generation.” or “You should put comfy furniture into your place. The feeling of being home is unforgettable”).
- **ethos**: appeals to the credibility established by personal experience/expertise (“I assure you the consequences of fracking are terrible. I have been living next to a pipeline since I was a child.” or “I assure you the consequences of fracking are terrible. I am a chemical engineer.”) as well as title/reputation (“I trust his predictions about climate change.

He is a Nobel Prize winner.” or “I trust his predictions about climate change. They say he is a very sincere person.”)

In operational terms, the workers were asked to select *true* for the persuasion mode used and *false* for the ones that were not applicable. They were given the choice to select from 1 to 3 modes for the same premise. If the workers did not select any modes, their HIT was rejected.

## 4 Annotation Results

The 78 discussion threads comprise 278 turns of dialogue consisting of 2615 propositions in 2148 total sentences. Of these sentences, 786 contain a claim and 1068 contain a premise. Overall at the sentence-level, 36.5% of sentences contain a claim and 49.7% contain a premise. 22% of sentences contain no annotations at all. In terms of claims,<sup>3</sup> 15.8% of sentences contain a rational evaluation, 8.7% contain an interpretation, and 7.3% contain an emotional evaluation, while only 2.5% contain agreement and 2.3% contain disagreement. For premises, 44% contain logos, 29% contain pathos, and only 3% contain ethos.

We computed Inter-Annotator Agreement for claims and premises by requiring 3 of the annotators to annotate an overlapping subset of 2 threads. We compare annotations at the sentence level, similar to previous work (Stab and Gurevych, 2014a), as most sentences contain only 1 proposition, making this approximation reasonable. We compute IAA using Krippendorff’s alpha (Krippendorff, 1970), obtaining 0.63 and 0.65, respectively. These scores are considered moderate agreement and are similar to the results on persuasive essays (Stab and Gurevych, 2014a).

We also compute IAA for types of premises, comparing the majority vote of the Turkers to gold labels from our most expert annotator (based on highest average pair-wise IAA). As Krippendorff’s alpha is calculated globally and compares each item directly between annotators, it is well-suited for handling the multi-label case here (Ravenscroft et al., 2016). The resulting IAA was 0.73.

Finally, we compute IAA for the types of claims, again comparing the majority vote to gold labels annotated by an expert linguist. The resulting IAA is 0.46, considered low agreement. This

<sup>3</sup>We took the majority vote among Turkers to determine the types of claims and premises.

result is in line with those attested in similar experiments (Walker et al., 2012).

In our case, we hypothesize that the nature of the claims provided as unit of annotations may have led to confusion. According to the expert linguist annotator, some of the claims are complex sentences being formed by two propositions liable to two different types of claims. In a sentence such as “Your first paragraph is intriguing, and I definitely agree with it,” for instance, the first proposition constitutes an emotional-evaluation, while the second an agreement. The choice of one of the two labels may, thus, give rise to divergent annotations.

#### 4.1 Qualitative analysis: the disagreement space

To investigate the disagreement space in the annotation of types of claims, we present a confusion matrix in Table 1 between the majority vote and the label chosen by each of the 5 Turkers. The major disagreement is between the claim types “interpretation” ( $C_I$ ) and “evaluation-rational” ( $C_{ER}$ ), followed by the pairs “evaluation-emotional” ( $C_{EE}$ )/ “evaluation-rational” ( $C_{ER}$ ). While the label “disagreement” ( $C_D$ ) also seems to be controversial, the scarcity of occurrences makes it less relevant for the analysis of the disagreement space. The higher consensus in the labeling of “agreement” ( $C_A$ ) versus other types of evaluations can be explained looking at linguistic triggers: while “agreement” is signaled by unambiguous linguistic clues (*I agree, you are right, yes*), the degree of rationality/emotions conveyed by a judgment is not always transparent given the semantics of the sentiment expressed, but may call for wider contextual features. Given a sentence such as “I don’t think I’m better than the people I’d be denying citizenship” it is clear that what the speaker is expressing is a subjective evaluation, while in the sentence “This is the best argument I have seen” the type of evaluation at stake depends on the criteria at the basis of the judgment.

In order to verify and explain difficulties encountered in deciding whether the claim is  $C_{ER}$  or  $C_I$  we compared the Turkers annotation with the gold annotations of an expert linguist annotator. The trends in the disagreement space are the same as those noticed among Turkers. The qualitative analysis shows that Turkers tend to misclassify interpretation ( $C_I$ ) as evaluation-rational ( $C_{ER}$ ). This is mainly due to a tendency of an-

| L \ M    | $C_A$ | $C_D$ | $C_{EE}$ | $C_{ER}$ | $C_I$ |
|----------|-------|-------|----------|----------|-------|
| $C_A$    | 186   | 8     | 17       | 35       | 19    |
| $C_D$    | 6     | 133   | 18       | 53       | 35    |
| $C_{EE}$ | 21    | 35    | 424      | 187      | 112   |
| $C_{ER}$ | 45    | 56    | 157      | 1150     | 220   |
| $C_I$    | 23    | 45    | 105      | 205      | 459   |

Table 1: Confusion Matrix for Claims  
L: individual labels M: majority vote

notating claims as evaluations in the presence of a sentiment word regardless of the overall meaning of the proposition: the sentence “The problem isn’t always bad parenting, though that can play a role, the problem is a black and white educational system” was annotated as an evaluation probably due to the axiological adjective *bad*. However, the primary meaning is not that of providing a negative judgment, but that of providing an explanation for a state of affairs (problems encountered at school).

## 5 Quantitative Analysis

In order to investigate what conceptual features are persuasive, we first observe correlations between types of argumentative components (premises and claims) as well as their position in the post. We then look at how different patterns are distributed in positive and/or negative threads.

### 5.1 Argumentative Components

We present an analysis of correlations between types of claims and premises, with the aim to check the presence of an ordering effect (research question 2). As we do not have supporting and attacking relations at this stage of the annotation process, we consider two approaches, both at the sentence-level, for analyzing dependencies.

We first report the results of the sequential transitions at the proposition level between types of claims (agreement, disagreement, rational evaluation, emotional evaluation, and interpretation) and premises (pathos, ethos, and logos, and their respective combinations). If the previous proposition is not labeled as claim or premise, we set the previous category to “None.” If the sentence is the start of a post, we set the previous category to “BOP” (beginning of post). We also include transitions to the end of the post (EOP). We present results for the annotations from the AMT workers in Figure 2. The heatmap represents the tran-

sition matrix, normalized by the row count. The rows represent the label for the previous proposition and the columns represent the label for the current proposition.

For the second approach, we report the counts for the type of premise given the most recent claim type in the post. We assume here that the premise always attaches to the preceding claim, providing an approximation for this type of structure. We chose this heuristic since we observed that users tend first to express their view and then back it up with subsequent arguments to achieve a clear argument structure as advocated by *ChangeMyView* submission rules. However, we acknowledge that premises may be positioned in front of a claim or refer anaphorically to a textually distant claim. We manually evaluated a sample of 100 premises-claims pairs: the correct pairs were identified 75% of the time. If the previous claim occurs either in the title or the previous post, we just indicate the previous claim to be “EOP.” This scenario occurs when the original poster writes a premise that depends on the main claim or when a post responds directly to a claim in a preceding post. The heatmap in Figure 3 represents the claim/premise distribution for AMT annotations, with claims as rows and premises as columns, normalized by the counts of premises.

We compute significance for individual cells using the chi-squared test for cells, computing a 2x2 contingency table. All results discussed have  $p < 0.001$  after the Bonferroni correction, unless otherwise specified. Considering only claims at the beginning of the post, rational evaluations (23%), agreements (5%), and interpretations (13%) are more likely to appear at the start than in general. On the other end, premises expressing *pathos* are less likely to appear at the end of the post (only 7% of the time), while less surprisingly, unannotated sentences (farewell messages, for example) are more likely to appear at the end (20% of the time). As far as sequences of modes of persuasion, arguments expressing *logos* or *pathos* are more likely to occur consecutively (for logos, 46% following logos and 48% following pathos and for pathos, 31% and 34% respectively) than in the overall distribution (37% *logos* and 24% *pathos*). Finally, *logos* is more likely to follow a rational evaluation (49% of the time) when compared to the overall distribution of *logos* and the same is true for emotional evaluations and pathos (39%).

As for premise/claim pairs, premises classified as *pathos* are in support of rational evaluations 34% of the time that pathos occurs, while *logos* supports rational evaluations 38% of the time ( $p < 0.05$ ) and *ethos* 28% of the time. Similarly, there is a slight preference ( $p < 0.05$ ) for pathos to support evaluation-emotional claims, with 20% of pathos arguments supporting that type, 17% of logos arguments and 17% of ethos supporting it, respectively. Finally, authors demonstrate a preference for *logos* when addressing the claims of an author in the previous post ( $p < 0.01$ ). The qualitative analysis of those cases reveals that when supporting rational evaluations, *pathos* arguments refer to situations that everyone could experience, as underlined by the use of the pronoun *you* in its impersonal use (e.g. “If you don’t break up, you are stuck with a person who doesn’t value you enough to stay loyal. It’s just a logical conclusion that breaking up is the right choice in most if not all situations.”).

## 5.2 Semantic types and persuasive role

To investigate whether certain types of claims/premises correlate with persuasive/non-persuasive messages (research question 3), we conduct a preliminary analysis of the relationship between claims and premises in different contexts, in winning vs. non winning arguments. We re-compute the transition matrix and conditional claim/premise matrix by splitting the dataset according to whether the responding poster received a delta or not. We also only consider the components written by the author of the response, and discard the posts from the original poster in order to understand whether certain patterns are more likely to be persuasive.

We compute statistical significance between the positive and negative label distributions and conditional and transition matrices using Pearson’s chi-squared test of independence. As the chi-squared test considers the distribution of the data and does not require equal sample sizes<sup>4</sup>, this test is appropriate for significance. We again use the Yates correction for low frequencies. For the AMT annotations, we obtain a p-value of  $p < 0.00001$  for all distributions: the unigram labels, the transition matrix, and the claim/premise matrix. For the gold annotations, the p-value of the overall label distri-

<sup>4</sup>Positive threads tend to be longer so they have more sentences and thus a higher number of claims and premises

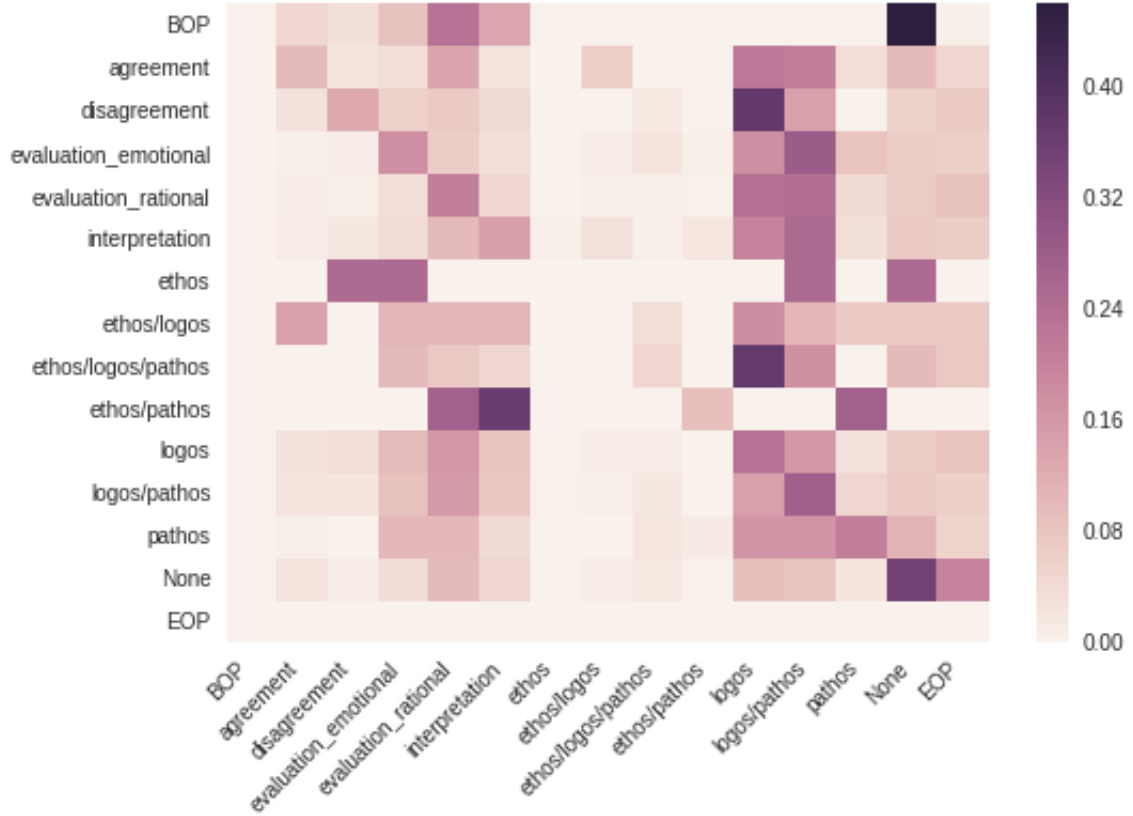


Figure 2: Transition Heatmap



Figure 3: Claim and Premise Heatmap

bution is  $p < 0.05$ , but for the transition matrix the p-value is  $p = 0.59$ , likely due to the very low counts for some cells. However, the value for the claim/premise matrix is  $p < 0.001$ , indicating significant differences even for this small dataset.

Finally, similar to the analysis of the entire dataset, we compute significance for individual cells using the same chi-squared test. We first find that for the unigram distribution rational evaluations are *less* likely to be found in winning arguments with 9% of propositions in positive and

14% in negative ( $p < 0.01$ ). When we consider the joint distribution of premise combinations, we find that *pathos* and *logos* are more likely to occur together in successful threads, with 23% and 17% respectively ( $p < 0.01$ ).

For the transition distribution, compared to positive threads, negative threads show fewer agreements opening up the posts ( $p < 0.05$ ). Agreeing with what was previously said by another speaker before expressing a possibly divergent opinion constitutes a traditional persuasive rhetorical strategy (Anscombe and Ducrot, 1983). In a sentence such as “I do agree that today’s moderates are potentially tomorrow’s conservatives. However this isn’t about being just a bit conservative. It’s about ...”, the speaker concedes the previous user’s point and then expresses a slightly contrasting point of view. In doing so, he exhibits his reasonableness and he avoids face-threatening disagreement. Moreover, positive threads are slightly more likely to show consecutive arguments of the same type (logos/logos; pathos/pathos) ( $p < 0.01$ ), suggesting the hypothesis that conceptual coherence plays a role as persuasive strategy. The reasons provided by the original posters for awarding a  $\Delta$  point frequently includes positive evaluations about the fol-

lowed reasoning lines (e.g. “Well thought out response”, “Thanks for the brilliant and well thought out answer”).

Examining premise/claim patterns qualitatively, it seems that positive threads generally feature more interpretations, especially based on arguments of the *logos* type, at the expense of the number of evaluations. This type of claim/premise pattern is likely to be perceived as less subjective.

Evaluations, even when of the rational type, necessarily contain a subjective component in assessing the criteria to judge something as more or less positive or negative: the judgment “networking is discriminatory” during the hiring process would not, for instance, be shared by someone who considers social skills as a crucial quality for a job candidate. On the other hand, interpretations, when backed up by *logos*, encode states of affairs presented as intersubjective (Nuyts, 2012). For instance, in the premise-claim pair “American patriots have a general mentality against immigration. This is prominent in many ads and political champagnes, namely the slogan ‘Creating jobs for americans’ ”, ads and political campaigns can be accessed by anyone. Since their goal is that of communicating a specific message to the public, the interpretation of their content promises to raise limited disagreement. This difference in degree of (inter)subjectivity is mirrored by the fact that evaluations, differently from interpretations, tend to be introduced by propositional attitude indicators at the first person singular (e.g. “I think”, “I find”, “I point out”) that put the speaker in a position of prominence as responsible for the truth of the asserted proposition. Moreover, evaluations are more frequently backed up by *pathos* arguments (e.g. the claim “Enjoying the moment is possible, but doesn’t make life have a point” and the matching premise “For once I die, all memories and all point is gone” (pathos).

## 6 Conclusions and Future Work

In this study we propose an annotation scheme for the identification of persuasive conceptual features. Compared to previous pilot works in the same vein, we distinguish different semantic types of premises and of claims with the long term goal of investigating their persuasive role. We empirically validate the devised procedure through a two-tiered annotation project on a sample of 78 threads from the subreddit *Change My View*.

While the annotation of argumentative components (claims, premises) was carried out by expert annotators, for the annotation of semantic types of premises and claims we relied on crowdsourcing. The annotation of premises and claims achieves moderate agreement, in line with state-of-the-art results. The same applies to the semantic types of premises, showing improvement with respect to previous attempts. The identification of the semantic types of claims appears to be more difficult due to the confusion between interpretations and rational evaluations. We plan to improve the guidelines to account for this difficulty.

In order to understand the persuasive role of the semantic types of claims and premises under study, we observe the recurrent combinations of argumentative components, their preferred position in the post and their distribution in winning and non winning threads.

Going forward, we plan to conduct a broader annotation project including the labeling of support/attack relations to be carried out as part of the identification of premise/claim pairs. We also plan to explore other aspects of the data. We expect that certain topics are more emotional or rational than others and winning arguments are generated accordingly. For example, moral issues may be more effective based on personal/emotional arguments while issues in science may require rational arguments. We also expect that the distribution of labels in the original post determines the effectiveness of a response, i.e. a post consisting mostly of emotional claims and pathos might require a similar response. Finally, we plan to experiment with predictive sequential models on claim and premise types and joint models for overall persuasiveness.

## Acknowledgement

This paper is based on work supported by DARPA-DEFT program. The second author has been supported by the Early Post Doc SNFS Grant n. *P2TIP1*<sub>165081</sub> and was mainly responsible for the design of the guidelines (sections 3.2./3.3.), supervision of annotations and the qualitative analysis of the results. The views expressed are those of the authors and do not reflect the official policy or position of the SNFS, Department of Defense or the U.S. Government. We would like to thank the annotators for their work and the anonymous reviewers for their valuable feedback.

## References

- Jean-Claude Anscombre and Oswald Ducrot. 1983. *L'argumentation dans la langue*. Editions Mardaga.
- Katie Atkinson and Trevor JM Bench-Capon. 2016. Argument schemes for reasoning about the actions of others. In *COMMA*, pages 71–82.
- Maria Becker, Alexis Palmer, and Anette Frank. 2016. Argumentative texts and clause types. *ACL 2016* page 21.
- Or Biran, Sara Rosenthal, Jacob Andreas, Kathleen McKeown, and Owen Rambow. 2012. Detecting influencers in written online conversations. In *Proceedings of the Second Workshop on Language in Social Media*. Association for Computational Linguistics, pages 37–45.
- Katarzyna Budzynska, Mathilde Janier, Chris Reed, Patrick Saint-Dizier, Manfred Stede, and Olena Yaskorska. 2014. A model for processing illocutionary structures and argumentation in debates. In *LREC*, pages 917–924.
- Katarzyna Budzynska and Chris Reed. 2011. Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. In *Computational Models of Natural Argument*.
- Robert B Cialdini. 2005. Influence: The psychology of persuasion collins. *Revised edition (October 7, 2005)*.
- Abhimanyu Das, Sreenivas Gollapudi, Emre Kıcıman, and Onur Varol. 2016. Information dissemination in heterogeneous-intent networks. In *Proceedings of the 8th ACM Conference on Web Science*. ACM, pages 259–268.
- Rory Duthie, Katarzyna Budzynska, and Chris Reed. 2016. Mining ethos in political debate. In *Proceedings of 6th International Conference on Computational Models of Argument (COMMA 2016)*. IOS Press, *Frontiers in Artificial Intelligence and Applications*.
- James B Freeman. 2000. What types of statements are there? *Argumentation* 14(2):135–157.
- Debanjan Ghosh, Aquila Khanam, Yubo Han, and Smaranda Muresan. 2016. Coarse-grained argumentation features for scoring persuasive essays. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 549–554.
- Ivan Habernal and Iryna Gurevych. 2016. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional lstm. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*.
- Colin Higgins and Robyn Walker. 2012. Ethos, logos, pathos: Strategies of persuasion in social/environmental reports. In *Accounting Forum*. Elsevier, volume 36, pages 194–208.
- Klaus Krippendorff. 1970. Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement* 30(1):61–70.
- John Lawrence and Chris Reed. 2016. Argument mining using argumentation scheme structures. In *COMMA*, pages 379–390.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5(1):1–167.
- Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion.
- Jan Nuyts. 2012. Notions of (inter) subjectivity. *English Text Construction* 5(1):53–76.
- Joonsuk Park, Cheryl Blake, and Claire Cardie. 2015. Toward machine-assisted participation in erulemaking: An argumentation model of evaluability. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*. ACM, pages 206–210.
- Andreas Peldszus and Manfred Stede. 2016. An annotated corpus of argumentative microtexts. In D. Mohammed and M. Lewinski, editors, *Argumentation and Reasoned Action - Proc. of the 1st European Conference on Argumentation, Lisbon, 2015*, College Publications, London.
- Chaim Perelman and Lucie Olbrechts-Tyteca. 1973. *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press.
- James Ravenscroft, Anika Oellrich, Shyamasree Saha, and Maria Liakata. 2016. [Multi-label annotation in scientific articles - the multi-label cancer risk assessment corpus](http://www.lrec-conf.org/proceedings/lrec2016/summaries/928.html). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*. <http://www.lrec-conf.org/proceedings/lrec2016/summaries/928.html>.
- Sara Rosenthal and Kathleen Mckeown. 2017. Detecting influencers in multiple online genres. *ACM Transactions on Internet Technology (TOIT)* 17(2):12.
- Christian Stab and Iryna Gurevych. 2014a. [Annotating argument components and relations in persuasive essays](#). In *Proceedings of COLING*

2014, *the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pages 1501–1510. <http://www.aclweb.org/anthology/C14-1142>.

Christian Stab and Iryna Gurevych. 2014b. Identifying argumentative discourse structures in persuasive essays. In *EMNLP*, pages 46–56.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pages 613–624.

Frans H van Eemeren and Frans Hendrik Eemeren. 2009. *Examining argumentation in context: Fifteen studies on strategic maneuvering*, volume 1. John Benjamins Publishing.

Marilyn A Walker, Jean E Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *LREC*, pages 812–817.

Zhongyu Wei, Yang Liu, and Yi Li. 2016. Is this post persuasive? ranking argumentative comments in the online forum. In *The 54th Annual Meeting of the Association for Computational Linguistics*, page 195.