RESPER : Computationally Modelling Resisting Strategies in Persuasive Conversations.

Ritam Dutt[*,1], **Sayan Sinha**[*,2], **Rishabh Joshi**[1], **Surya Shekhar Chakraborty**[3], **Meredith Riggs**[1], **Xinru Yan**[1], **Haogang Bao**[1], **Carolyn Penstein Rosé**[1]

[1]Carnegie Mellon University, [2]Indian Institute of Technology Kharagpur, [3]Zendrive Inc
{rdutt,rjoshi2,mriggs,xinruyan,haogangb,cprose}@cs.cmu.edu,
sayan.sinha@iitkgp.ac.in, suryaschak@gmail.com

Abstract

Modelling persuasion strategies as predictors of task outcome has several real-world applications and has received considerable attention from the computational linguistics community. However, previous research has failed to account for the resisting strategies employed by an individual to foil such persuasion attempts. Grounded in prior literature in cognitive and social psychology, we propose a generalised framework for identifying resisting strategies in persuasive conversations. We instantiate our framework on two distinct datasets comprising persuasion and negotiation conversations. We also leverage a hierarchical sequence-labelling neural architecture to infer the aforementioned resisting strategies automatically. Our experiments reveal the asymmetry of power roles in non-collaborative goal-directed conversations and the benefits accrued from incorporating resisting strategies on the final conversation outcome. We also investigate the role of different resisting strategies on the conversation outcome and glean insights that corroborate with past findings. We also make the code and the dataset of this work publicly available at https://github.com/americast/resper.

1 Introduction

Persuasion is pervasive in everyday human interactions. People are often exposed to scenarios that challenge their existing beliefs and opinions, such as medical advice, election campaigns, and advertisements (Knobloch-Westerwick and Meng, 2009; Bartels, 2006; Speck and Elliott, 1997). Of late, huge strides have been taken by the Computational Linguistics community to advance research in persuasion. Some seminal works include identifying persuasive strategies in text (Yang et al., 2019) and conversations (Wang et al., 2019), investigating the interplay of language and prior beliefs on successful persuasion attempts (Durmus and Cardie, 2018; Longpre et al., 2019), and generating persuasive dialogues (Munigala et al., 2018).

However, a relatively unexplored domain by the community is the investigation of resisting strategies employed to foil persuasion attempts. As succinctly observed by Miller (1965): "In our daily lives we are struck not by the ease of producing attitude change but by the rarity of it." Several works in cognitive and social psychology (Fransen et al., 2015a; Zuwerink Jacks and Cameron, 2003) have put forward different resisting strategies and the motivations for the same. However, so far, there has not been any attempt to operationalise these strategies from a computational standpoint. We attempt to bridge this gap in our work.

We propose a generalised framework, grounded in cognitive psychology literature, for automatically identifying resisting strategies in persuasion oriented discussions. We instantiate our framework on two publicly available datasets comprising persuasion and negotiation conversations to create an annotated corpus of resisting strategies.

Furthermore, we design a hierarchical sequence modelling framework, that leverages the conversational context to identify resisting strategies automatically. Our model significantly outperforms several neural baselines, achieving a competitive macro-F1 score of 0.56 and 0.66 on the persuasion and negotiation dataset, respectively.

We refer to our model as RESPER, which is not only an acronym for **Res**isting **Per**suasion, but also a play on the word ESPer: a person with extrasensory abilities. The name is apt since we observe that incorporating such resisting strategies could provide additional insight on the outcome of the conversation. In fact, our experiments reveal that the resisting strategies are better predictors of conversation success for the persuasion dataset than the

[*] denotes equal contribution

Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, pages 78–90
April 19 - 23, 2021. ©2021 Association for Computational Linguistics

strategies employed by the persuader. We also observe that the buyer's strategies are more influential in negotiating the final price. Our findings highlight the asymmetric nature of power roles arising in non-collaborative dialogue scenarios and form motivation for this work.

2 Related Works

The use of persuasion strategies to change a person's view or achieve a desired outcome finds several real-world applications, such as in election campaigns (Knobloch-Westerwick and Meng, 2009; Bartels, 2006), advertisements (Speck and Elliott, 1997), and mediation (Cooley, 1993). Consequently, several seminal NLP research have focused on operationalising and automatically identifying persuasion strategies (Wang et al., 2019), propaganda techniques (Da San Martino et al., 2019), and negotiation tactics (Zhou et al., 2019), as well as the impact of such strategies on the outcome of a task (Yang et al., 2019; He et al., 2018; Joshi et al., 2021). However, there is still a dearth of research from a computational linguistic perspective investigating resisting strategies to foil persuasion.

Resisting strategies have been widely discussed in literature from various aspects such as marketing (Heath et al., 2017), cognitive psychology (Zuwerink Jacks and Cameron, 2003), and political communication (Fransen et al., 2015b) . Some notable works include the identification and motivation of commonly-used resisting strategies (Fransen et al., 2015a; Zuwerink Jacks and Cameron, 2003), the use of psychological metrics to predict resistance (San José, 2019; Ahluwalia, 2000), and the design of a framework to measure the impact of resistance (Tormala, 2008). However, these works have mostly relied on qualitative methods, unlike ours, which adopts a data-driven approach. We propose a generalised framework to characterise resisting strategies and employ state-of-the-art neural models to infer them automatically. Thus our work can be considered complementary to past research.

The closest semblance to our work in NLP literature ties in with argumentation, be it essays (Carlile et al., 2018), debates (Cano-Basave and He, 2016), or discussions on social media platforms (Al-Khatib et al., 2018; Zeng et al., 2020). Such works have revolved mostly on analysing argumentative strategies and their effect on others.

Recently, Al Khatib et al. (2020) demonstrated that incorporating the personality traits of the resistor was influential in determining their resistance to persuasion. Such an observation acknowledges the power vested in an individual to resist change to their existing beliefs. Our work exhibits significant departure from this because we explicitly characterise the resisting strategies employed by the user. Moreover, our work focuses on the general domain of non-collaborative task-oriented dialogues, where several non-factual resisting strategies are observed, making it distinctly different from argumentation (Galitsky et al., 2018). We assert that focusing on both parties is imperative to get a complete picture of persuasive conversations.

3 Framework

In this section, we describe the datasets, the resisting strategies employed, and the annotation framework to instantiate the strategies.

3.1 Dataset Employed

We choose persuasion-oriented conversations, rather than essays or advertisements (Yang et al., 2019), since we can observe how the participants respond to the persuasion attempts in real-time. To that end, we leverage two publicly available corpora on persuasion (Wang et al., 2019) and negotiation (He et al., 2018). We refer to these datasets as "Persuasion4Good" or P4G and "Craigslist Bargain" or CB hereafter.

P4G comprises conversational exchanges between two anonymous Amazon Mechanical Turk workers with designated roles of the persuader, ER and persuadee, EE. ER had to convince EE to donate a part of their task earnings to the charity *Save the Children*. We investigate the resisting strategies employed only by EE in response to the donation efforts. We emphasise that the conversational exchanges are not scripted, and the task is set up so that a part of EE's earnings is deducted if they agree to donate. Since there is a monetary loss at stake for EE, we expect them to resist.

CB consists of simulated conversations between a buyer (BU) and a seller (SE) over an online exchange platform. Both are given their respective target prices and employ resisting strategies to negotiate the offer.

We choose these datasets since they involve non-collaborative goal-oriented dialogues. As a result, we can definitively assess the impact of different resisting strategies on the goal.

Resisting Strategy	Persuasion (P4G)	Negotiation (CB)
Source Derogation	Attacks/doubts the organisation's credibility. *My money probably won't go to the right place*	Attacks the other party or questions the item. *Was it new denim, or were they someone's funky old worn out jeans?*
Counter Argument	Argues that the responsibility of donation is not on them or refutes a previous statement. *There are other people who are richer*	Provides a non-personal argument/factual response to refute a previous claim or to justify a new claim. *It may be old, but it runs great. Has lower mileage and a clean title.*
Personal Choice	Attempts to saves face by asserting their personal preference such as their choice of charity and their choice of donation. *I prefer to volunteer my time*	Provides a personal reason for disagreeing with the current situation or chooses to agree with the situation provided some specific condition is met. *I will take it for $300 if you throw in that printer too.*
Information Inquiry	Ask for factual information about the organisation for clarification or as an attempt to stall. *What percentage of the money goes to the children?*	Requests for clarification or asks additional information about the item or situation. *Can you still fit it in your pocket with the case on?*
Self Pity	Provides a self-centred reason for not being able/willing to donate at the moment. *I have my own children*	Provides a reason (meant to elicit sympathy) for disagreeing with the current terms. *$130 please I only have $130 in my budget this month.*
Hesitance	Attempts to stall the conversation by either stating they would donate later or is currently unsure about donating. *Yes, I might have to wait until my check arrives.*	Stalls for time and is hesitant to commit; specifically, they seek to further the conversation and provide a chance for the other party to make a better offer. *Ok, would you be willing to take $50 for it?*
Self-assertion	Explicitly refuses to donate without even providing a factual/personal reason *Not today*	Asserts a new claim or refutes a previous claim with an air of finality/ confidence. *That is way too little.*

Table 1: Framework describing the resisting strategies for persuasion (P4G) and negotiation (CB) datasets. We emphasise that Information Inquiry is not a resisting strategy for CB. Examples of each strategy are italicised.

Properties	**P4G**	**CB**
# of conversations	530	800
Max # of utterances/conversation	76	44
Avg # of utterances/conversation	36.34	11.94
Max # of tokens/utterance	90	93
Avg # of tokens/utterance	11.03	14.62
Vocabulary size	6137	5370

Table 2: Description for the Persuasion (P4G) (Wang et al., 2019) and Negotiation (CB) (He et al., 2018) datasets

3.2 Framework Description

In this subsection, we briefly describe the resisting strategies commonly referenced in social and cognitive psychology literature. This enables us to design a unified framework for the two datasets, built upon common underlying semantic themes. Fransen et al. (2015a) identified 4 major clusters of resisting strategies, namely **contesting** (Wright, 1975; Zuwerink Jacks and Cameron, 2003; Abelson and Miller, 1967), **empowerment** (Zuwerink Jacks and Cameron, 2003; Sherman and Gorkin, 1980), **biased processing** (Ahluwalia, 2000), and **avoidance** (Speck and Elliott, 1997). Each individual category can be subdivided into finer categories showcased in italics henceforth.

Contesting refers to attacking either the source of the message (*Source Derogation*) or its content (*Counter Argumentation*). A milder form of contesting involves seeking clarification or information termed *Information Inquiry*. Prior work has shown a positive association between working knowledge and one's ability to resist persuasion (Wood and Kallgren, 1988; Luttrell and Sawicki, 2020). Therefore, *Information Inquiry* can be interpreted as a form of resistance where the resistor seeks to satisfy their doubts because they are sceptical of the persuader's intents or messages. This is prominent in certain conversations in P4G where a sceptical EE questions the charity's legitimacy.

Empowerment strategies encompass reinforcing one's personal preference to refute a claim (*Attitude Bolstering*) (Sherman and Gorkin, 1980), attempting to arouse guilt in the opposing party (*Self Pity*) (Vangelisti et al., 1991; O'Keefe, 2002), stating one's wants outright (*Self Assertion*) (Zuwerink Jacks and Cameron, 2003), or seeking validation from like-minded people (*Social Validation*) (Fransen et al., 2015a). Overall, empowerment strategies drive the discussion towards the resistor's self as opposed to attacking the persuader.

Biased processing mitigates external persuasion by selectively processing information that conforms with one's opinion or beliefs (Fransen et al., 2015a). For simplicity, we subsume strategies that denote personal preference, namely *Attitude Bolstering* and *Biased Processing*, into a unified category *Personal Choice*. We refrain from incorporating *Self Assertion* into the *Personal Choice* category since it deals with bolstering one's confidence and not one's opinions or attitudes. The subtle difference is highlighted in Table 1.

Avoidance strategies distance the resistor from persuasion, either physically or mechanically, or refuse to engage in topics that induce cognitive dissonance (Fransen et al., 2015a). However, in the context of task-oriented conversations, wherein participants are expected to further a goal, avoidance often manifests as *Hesitance* to commit to the current situation.

We identify seven major resisting strategies across the datasets, namely *Source Derogation, Counter Argumentation, Information Inquiry, Personal Choice, Self Pity, Hesitance,* and *Self Assertion*. Since the datasets comprise two-party conversations between strangers, *Social Validation*, which requires garnering the support of others, was absent. We now describe how these resisting strategies were instantiated in the following section.

3.3 Instantiating the Resistance Framework

We emphasise that although the description and meaning of a strategy remain the same across the two datasets, their semantic interpretation depends on the context. For example, scepticism towards the charity in P4G and criticism of the product in CB are instances of *Source Derogation*. This is because ER represents the charity, whereas the seller is being accused of selling an inferior product. Likewise, we instantiate the predicates for the remaining six resisting strategies for the two datasets, with examples in Table 1.

We label the utterances of persuadee (EE) in P4G and the buyers (BU) and sellers (SE) in CB with at least one of the seven corresponding resisting strategies, or 'Not-A-Strategy' if none applies. The 'Not-A-Strategy' label includes greetings, off-task discussions, agreement, compliments, or other tokens of approval. We acknowledge that an utterance can have more than one resisting strategy embedded in it. For example, the utterance "The price is slightly high for used couches, would you come down to 240 if I also picked them up?", is an instance of both *Personal Choice* and *Counter Argumentation*.

We also note that ***Information-Inquiry* is not a resisting strategy for CB** since asking additional information/clarification is an expected behaviour before finalising a deal. We keep the label nevertheless to show comparison with P4G. We present the flowchart detailing the annotation framework in Figure 3 of Appendix.

3.4 Annotation Procedure and Validation

We describe the annotation procedure for both the CB and P4G dataset here and its subsequent validation. For CB, three authors independently annotated five random conversations adhering to the flowchart. If the conversations chosen were simple or had few labels, a new set of 5 conversations were taken up. This constitutes one round. After each round, the Fleiss Kappa score was computed, and the authors discussed to resolve the disagreements and revise the flowchart. Then began the next round on a new set of 5 random conversations. For CB, 5 rounds of revision were carried out over 24 conversations, until a high Fleiss kappa (0.790) (Fleiss, 1971) was obtained. Finally, the three authors independently went ahead and annotated approximately 250 distinct conversations, yielding a corpus of 800 CB conversations. Our annotation procedure requires a rigorous reliable refinement phase but a comparatively faster annotation phase by dividing the annotation between the authors. Thus the conversations annotated by each author were mutually exclusive. Similarly, for P4G dataset, four authors annotated 3 conversations per round, since a conversation in P4G was comparatively longer. 4 rounds of revision across 12 conversations was done to achieve the final kappa-score of 0.787. The four authors then went ahead and divided the task of annotating the 500 conversations amongst themselves. We show an annotated conversation snippet for the two datasets in Table 3.

3.5 Dataset Statistics

The P4G and CB datasets comprise 530 and 800 labelled conversations, respectively, spanning an average of 37 and 12 utterances per conversation.

Role	Text	Strategy
	Negotiation (CB)	
SE	I have a wonderful phone for you if you are interested.	No Strategy
BU	I am interested. Did you just buy it?	Info inquiry
SE	I bought it two weeks ago but it just wasn't what I needed anymore.	No Strategy
BU	Would you be willing to work with the price?	Hesitance
SE	Yes we can negotiate.	No Strategy
BU	If I come today would you accept $56 I can bring it now?	Per Choice
SE	How about 65 and I can deliver it to you now?	Per Choice
BU	Can you go $60 Kind of all I have right now ?	Self Pity
SE	Yes I can.	No Strategy
	Persuasion (P4G)	
ER	Hello, Save the Children looks like an interesting organisation.	-
EE	i would like to know more about it	Info Inquiry
..	..	..
EE	thanks i will definitely check it out	Hesitance
ER	They also promote children's rights and provide relief when needed.	-
EE	and where does the money go if i do donate ?	Info Inquiry
EE	Straight to the organisation?	Info Inquiry
ER	Yes, it goes straight to the organisation, where it can be used to help many children.	-
EE	because some organisations do not divide the money properly	Source Derogation
ER	This organisation has been checked by some groups, and they divide the money properly.	-
..	..	..
EE	I will certainly consider it	No Strategy

Table 3: Examples of annotation snippets for the Persuasion (P4G) and Negotiation (CB). The utterances of the EE and the SE are highlighted in cyan. Some strategies are shortened, like Info Inquiry, and Per Choice for Information Inquiry and Personal Choice.

The datasets cover two distinct persuasion scenarios and also illustrate the rights and obligations shown by the participants. For example, in P4G, EE comes into the interaction blind and is unaware of the donation attempt. We encounter several conversations where EE is willing to donate since it resonates with their beliefs, and no resisting strategies are observed. However, for CB, the participants received prior instructions to negotiate a deal, and hence resisting strategies were more prominent. We present the frequency distribution of the seven strategies in Table 4. We observe that the distributions of strategies are skewed for both the datasets and is more pronounced for P4G, where 'Not-A-Strategy' accounts for the lion's share. We also see that the buyer exhibits more resisting strategies than the seller highlighting the asymmetric role of the two participants.

Nevertheless, we reiterate that the resisting strategies we propose are applicable for both the domains. In the next section, we propose the framework to infer such strategies automatically.

Strategy	Persuasion (P4G)	Negotiation (CB)	
	EE	BU	SE
Source Derogation	2.16	7.61	0.44
Counter Argument	2.28	3.74	6.06
Personal Choice	2.52	9.43	8.49
Information Inquiry	7.19	18.27	0.38
Self Pity	1.58	4.66	0.34
Hesitance	1.76	15.78	9.14
Self-assertion	0.94	2.20	5.05
Not a strategy	81.56	38.30	70.09

Table 4: Proportion of resisting strategies (in %)for the Persuasion (P4G) and Negotiation (CB) dataset. The strategies are observed only for the persuadee (EE) in P4G and for both buyer (BU) and seller (SE) in CB.

4 Methodology

In this section, we describe the methodology adopted for inferring the resisting strategies in persuasion dialogues and how they can be leveraged to determine the dialogue's outcome.

4.1 Resisting Strategy prediction

We model the task of identifying resisting strategies as a sequence labelling task. We assign each utterance in the dialogues with a label representing either one of the seven resisting strategies or *Not-A-Strategy*.

Since the resisting strategies, by definition, occur in response to the persuasion attempts, our model architecture needs to be cognizant of the conversational history. To that end, we adopt a hierarchical neural network architecture, similar to Jiao et al. (2019), to infer the corresponding resisting strategy. The architecture leverages the previous conversational context in addition to the current contextualised utterance embedding. Our choice is motivated by the recent successes of hierarchical sequence labelling frameworks in achieving state-of-the-art performance on several dialogue-oriented tasks. Some myriad examples include emotion recognition (Majumder et al., 2019; Jiao et al., 2019), dialogue act classification (Chen et al., 2018; Raheja and Tetreault, 2019), face act prediction (Dutt et al., 2020), open domain chit-chat (Zhang et al., 2018; Kumar et al., 2020) and the like. We hereby adopt this as the foundation architecture for our work and refer to our instantiation of the architecture as ResPer.

Architecture of ResPer: An utterance u_j

We acknowledge that an utterance can have multiple labels. However, such utterances comprise only 1.2% and 3.85% of the P4G and the CB datasets, respectively. In such cases, the label is randomly selected.

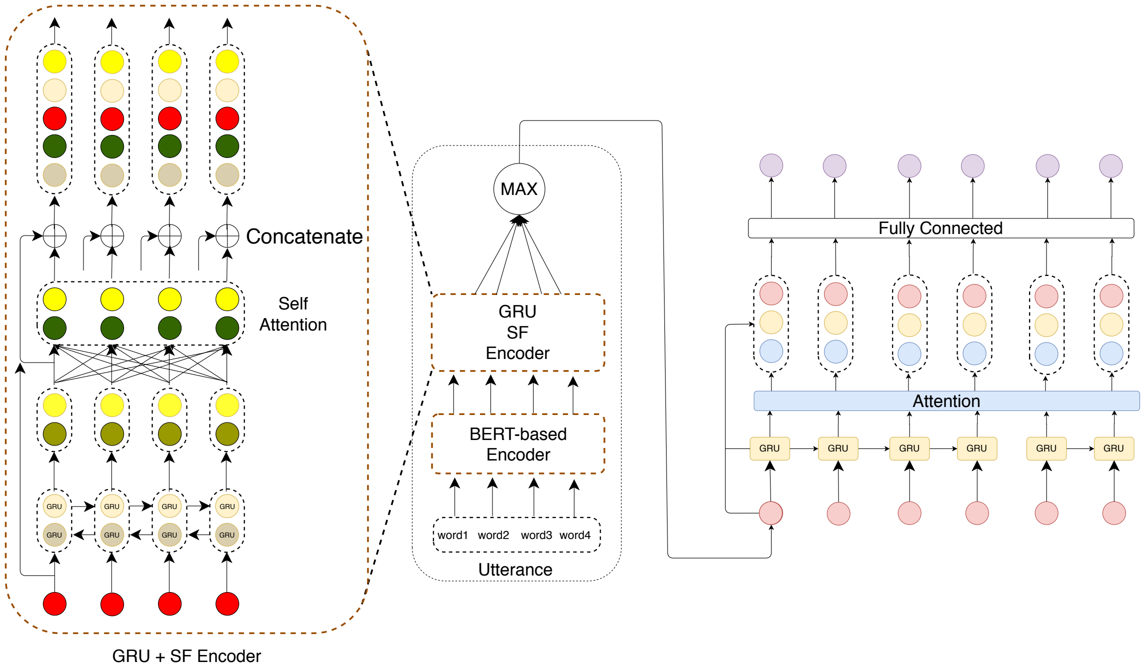

Figure 1: A diagram illustrating how RESPER works. The encoder shown on the left takes the BERT representations of a token as input and passes it through a BiGRU layer followed by Self Attention. The outputs from BERT, BiGRU and self-attention are then concatenated to form the output. Max-pooling over this output yields the corresponding utterance embedding. This utterance representation is passed through a uni-directional GRU followed by Masked-Self-Attention and fusion to yield the contextualised utterance embedding.

is composed of tokens $[w_0, w_1, ..., w_K]$ represented by their corresponding embeddings $[e(w_0), e(w_1), ..., e(w_K)]$. In RESPER, we obtain these using a pre-trained BERT model (Devlin et al., 2019). We pass these contextualised word representations through a bidirectional GRU to obtain the forward $\overrightarrow{h_k}$ and backward $\overleftarrow{h_k}$ hidden states of each word, before passing them into a Self-Attention layer. This gives us the corresponding attention outputs, $\overrightarrow{ah_k}$ and $\overleftarrow{ah_k}$ as described below.

$$\overrightarrow{h_k} = \mathrm{GRU}\left(e\left(w_k\right), \overrightarrow{h_{k-1}}\right)$$
$$\overleftarrow{h_k} = \mathrm{GRU}\left(e\left(w_k\right), \overleftarrow{h_{k+1}}\right)$$
$$\overrightarrow{ah_k} = SelfAttention(\overrightarrow{h_k})$$
$$\overleftarrow{ah_k} = SelfAttention(\overleftarrow{h_k})$$

Finally, we concatenate the contextualised word embedding with the GRU hidden states and Attention outputs in the *fusion layer* to obtain the final representation of the word $e_c(w_k)$. We represent the bias as b_w. Here, We perform max-pooling over the fused word embeddings to obtain the j^{th} utterance embedding, $e(u_j)$.

$$e_c(w_k) = \tanh(W_w[\overrightarrow{ah_k}; \overrightarrow{h_k}; e(w_k); \overleftarrow{h_k}; \overleftarrow{ah_k}] + b_w)$$
$$e(u_j) = \max(e_c(w_1), e_c(w_2), ...e_c(w_K))$$

We use a unidirectional GRU and Masked Self-Attention to encode conversational context, to ensure that the prediction for the j^{th} utterance is not influenced by future utterances. Similarly, we calculate the contextualized representation of an utterance $e_c(u_j)$ using the conversation context. We pass $e(u_j)$ through a uni-directional GRU that yields the forward hidden state $\overrightarrow{H_j}$. Masked Self-Attention over the previous hidden states, yields $\overrightarrow{AH_j}$. We fuse $e(u_j)$, $\overrightarrow{H_j}$ and $\overrightarrow{AH_j}$ before passing it through a linear layer with tanh activation to obtain $e_c(u_j)$.

We project the final contextualised utterance embedding $e_c(u_j)$ onto the state space of resisting strategies. We apply softmax to obtain a probability distribution over the strategies, with Negative Log-Likelihood (NLL) as the loss function to obtain the strategy loss.

4.2 Conversation Outcome prediction

We further investigate the impact of resisting strategies on the outcome of the conversation. We represent a strategy as a fixed dimensional embedding initialised at random. We subsequently encode a sequence of strategies by passing them through a unidirectional GRU to obtain a final representation for the sequence. We project the representation onto a binary vector which encodes for the conversation outcome. We apply softmax with NLL across all the conversations to obtain the outcome prediction loss.

5 Experiments

In this section, we describe the baselines and evaluation metrics. We present the experimental details of our model in Table 5.

5.1 Baselines

Resisting strategy prediction: We experiment with standard neural baselines for text classification, which have also been used in classifying persuasion strategies, namely CNN (Kim, 2014; Wang et al., 2019) and BiGRU (Yang et al., 2019). To ensure a fair comparison, we introduce pre-trained BERT-embeddings (Devlin et al., 2019) as input to the baselines, henceforth denoted as BERT-CNN and BERT-BiGRU. Furthermore, to inspect the impact of conversational history, we remove the conversational GRU from RESPER such that the utterance embedding $e(u_j)$ is directly used for prediction. We refer to this architecture as BERT-BiGRU-sf, since it employs self-attention(s) and fusion (f) on top of BERT-BiGRU. Finally, we experiment with the best performing HiGRU-sf model of Jiao et al. (2019) as another baseline.
Conversation success prediction: The notion of conversation success depends on the choice of dataset. For P4G, we consider the resisting strategies to be successful if the persuadee (EE) refused to donate to charity. For CB, we adopt the same notion of success as Zhou et al. (2019), namely when the seller (SE) can sell at a price greater than the median sale-to-list ratio r.

$$r = \frac{\text{sale price} - \text{buyer target price}}{\text{listed price} - \text{buyer target price}} \quad (1)$$

To observe the effect of conversation success, we experiment with strategies of both the parties involved. For P4G, we encode separately (i) the persuasion strategies of ER as identified by Wang et al. (2019), (ii) the resisting strategies employed by EE and (iii) both the persuasion and resisting strategies. Likewise, for CB, we encode the resisting strategies of only (i) the buyer (BU) (ii) the seller (SE) (iii) both. These experiments would enable us to investigate which party has a greater influence on conversation success.

Hyper-parameter	Search space	Final Value
learning-rate (lr)	1e-3 to 1e-5	1e-4
Batch-size	-	1 conversation
#Epochs	< 100	30.8, 22
lr-decay	-	0.5 every 20 epochs
d_{h1}	-	1024
d_{h2}	-	300

Table 5: Here we describe the search-space of all the hyper-parameters used in our experiments and describe the search space we used to find the hyper-parameters. d_{h1}, d_{h2} represents the hidden dimensions of the Utterance GRU and the Conversation GRU.

5.2 Evaluation metrics

We adopt the same evaluation procedure for both the resisting strategy and the conversation outcome prediction task across the datasets. In either case, we perform five-fold cross-validation due to paucity of annotated data. We report performance in terms of the weighted and macro F1-scores across the five folds. Our choice of the metric is motivated by the high label imbalance, as observed in Table 4.

6 Results

In this section, we answer the following :

Q1. How well does RESPER identify resisting strategies for Persuasion and Negotiation?

Q2. Are resisting strategies good predictors of conversation success? What insights can one glean from the results?

6.1 Predicting resisting strategies

We present the results for the automated identification of resisting strategies in Table 6. We observe that all the models achieve a comparatively lower performance on P4G, mainly due to the higher proportion of 'Not-a-Strategy' labels for the latter. We gauge the benefits of incorporating conversational context by the significant improvement of Macro F1 score by 0.036 and 0.011 for P4G and CB respectively. In fact, RESPER outperforms all the proposed baselines significantly.
Error Analysis: We present the confusion matrix for predicting resisting strategies using RESPER on the Persuasion (P4G) and Negotiation (CB)

Weighted F1 Scores are calculated by taking the average of the F1 scores for each label weighted by the number of true instances for each label.

We estimate the statistical significance using the paired bootstrapped test of Berg-Kirkpatrick et al. (2012), due to the small number of data (Dror et al., 2018).

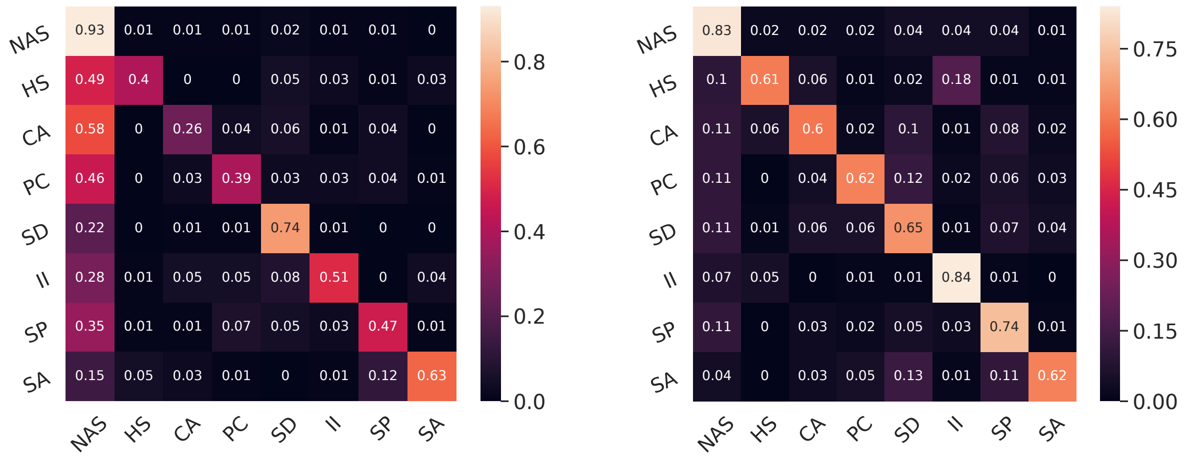

Figure 2: Confusion matrix for resisting strategies for the Persuasion (P4G) and Negotiation (CB) datasets on the left and right respectively. Each resisting strategy is represented as its initial (**S**elf **P**ity) as SP. True and Predicted Labels have been plotted on the X-axis and the Y-axis respectively.

Model	Persuasion (P4G)		Negotiation (CB)	
	M-F1	W-F1	M-F1	W-F1
CNN	0.261	0.757	0.560	0.706
BERT + CNN	0.508	0.819	0.651	0.751
HiGRU-sf	0.446	0.788	0.605	0.734
BERT + BiGRU	0.514	0.815	0.647	0.747
BERT + BiGRU-sf	0.522	0.814	0.649	0.750
RESPER	**0.558**	**0.828**	**0.662**	**0.767**

Table 6: Results of RESPER and other baselines on the resistance strategy prediction task on the Persuasion and CB dataset. The metrics used for evaluation are Macro F1 and Weighted F1 represented as M-F1 and W-F1 respectively. The best results are in bold.

datasets in Figures 2(a) and 2(b) respectively. We observe that most classification errors occur when a resisting strategy is incorrectly inferred as 'Not-A-Strategy'. The effect is more prevalent for P4G since 'Not-A-Strategy' comprises 80% of all annotated labels. Other notable instances of misclassification for P4G occurs when *Self Assertion* is predicted as *Self Pity* since both strategies refer to one's self. These strategies occur so infrequently (see Table 4) that the models lack sufficient information to distinguish between the two categories. Likewise, for the CB corpus, *Hesitance* utterances which constitute a price request, are often posed as questions. This causes the model to predict the strategy as *Information Inquiry* instead. *Self Assertion* is often incorrectly marked as *Source Derogation* possibly because it often takes a firm stance, and is likely to disparage the other party in the process, thereby confusing the model.

Persuasion (P4G)			Negotiation (CB)		
User	Macro-F1	W-F1	User	Macro-F1	W-F1
ER	0.588	0.620	BU	**0.618**	**0.640**
EE	0.618	0.640	SE	0.462	0.508
Both	**0.646**	**0.671**	Both	0.605	0.626

Table 7: We observe the impact of incorporating sequence of strategies on conversation outcome prediction in terms of Macro-F1 and Weighted-F1 score. For P4G, we observe strategies of the persuader (ER), persuadee (EE) and both. For CB, we observe strategies of the buyer (BU), seller (SE) and both.

6.2 Conversation Outcome Prediction

We observe how the sequence of strategies adopted by the two participants have a disproportionate impact on the final conversation outcome in Table 7. It is interesting to note that the resisting strategies for the persuadee have a greater effect on the conversation outcome (macro-F1 score of 0.62) than the persuasion strategies themselves (macro-F1 score of 0.59). Moreover, incorporating both the persuasion and resisting strategies boosts the prediction performance even further to 0.65.

We also observe an asymmetry in the roles of the buyer (BU) and the seller (SE) for the CB dataset. We observe that BU's strategies are significantly more effective in deciding the conversation outcome, probably because buyers demonstrate a higher number of resisting strategies. These experiments highlight the importance of incorporating resisting strategies to gain a complete picture.

6.3 Comparative Analysis of Strategies

Emboldened by the success of resisting strategies to infer the conversational outcome, we probe deeper to investigate the impact of individual strategies. We apply logistic regression with the frequency of strategies, of either participant, as the features while the outcome variable denotes conversation success. We observe the coefficients of the strategies to infer their correlation with conversation success and their corresponding p-values to determine whether the correlation was indeed statistically significant. Our procedure follows previous work in identifying influential persuasion strategies (Yang et al., 2019; Wang et al., 2019). We present the results of this analysis in Table 8.

	Persuasion (P4G)	Negotiation (CB)	
Strategy	EE	BU	SE
Not-A-Strategy	-0.008	0.287**	-0.138
Hesitance	0.344	0.328*	0.266
Counter Argument	-0.014	-0.256	0.429*
Personal Choice	0.153	0.126	0.164
Information Inquiry	0.180*	0.091	-0.704
Source Derogation	0.043	0.052	-0.455
Self Pity	0.103	0.081	-0.314
Self Assertion	0.843*	-0.576*	-0.040

Table 8: Coefficients of the different persuasion strategies corresponding to the persuadee, EE in Persuasion and the buyer, BU, and seller, SE in Negotiation. A value of * and ** means the strategy is signficant with p-value ≤ 0.05 and 0.01 respectively.

For P4G, all the resisting strategies for persuasion apart from *Counter-Argumentation* are positively correlated with a refusal to donate. The highest impact stems from *Self Assertion*. Previous research (Fransen et al., 2015a; Zuwerink Jacks and Cameron, 2003) has noticed that *Self Assertion* is prominent amongst individuals with high self-esteem. Such individuals are confident about their beliefs and less likely to conform. Similarly, a high positive coefficient for *Information Inquiry* can be attributed as follows. EE inquires information about the charity not only as a means to verify their legitimacy, but also to gain the knowledge they can exploit to their advantage. An innocuous question like ‘Where will my money go?’ would enable EE to assert that they are keener to help children in their own country instead, thereby resisting the donation attempt and saving face.

The CB scenario setup ensures that the coefficients of the strategies set for BU and SE would be anti-correlated, which holds for the Table 8. Like P4G, a high negative coefficient of *Self Assertion* signifies that SE’s price is disagreeable to BU - they would instead not buy. Moreover, the high coefficient of *Counter Argumentation* justifies that it is an effective tactic for both parties.

7 Conclusion

We present a generalised computational framework grounded in cognitive psychology to operationalise resisting strategies employed to counter persuasion. We identify seven distinct resisting strategies that we instantiate on two publicly available corpora comprising persuasion and negotiation conversations. We adopt a hierarchical sequence labelling architecture to infer the resisting strategies automatically and observe that our model achieves competitive performance for both datasets. Furthermore, we examine the interplay of resisting strategies in determining the final conversation outcome, which corroborates with previous findings. In the future, we would like to explore better models to encode the strategy information and apply our framework to improve personalised persuasion and negotiation dialogue systems. We would also like to study the influence of other confounding factors such as power dynamics on the outcomes of conversations featuring resisting strategies.

Acknowledgments

We thank the anonymous EACL reviewers for their insightful comments and constructive feedback. This research was funded in part by NSF Grants (IIS 1917668 and IIS 1822831) and Dow Chemical. The first author would also like to acknowledge his best friend, Ahana Sadhu, for her constant support and motivation, who unfortunately and untimely left us this year.

References

Robert P Abelson and James C Miller. 1967. Negative persuasion via personal insult. *Journal of Experimental Social Psychology*, 3(4):321–333.

Rohini Ahluwalia. 2000. Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research*, 27(2):217–232.

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. Exploiting personal characteristics of debaters for predicting persuasiveness. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7067–7072, Online. Association for Computational Linguistics.

Khalid Al-Khatib, Henning Wachsmuth, Kevin Lang, Jakob Herpel, Matthias Hagen, and Benno Stein. 2018. Modeling deliberative argumentation strategies on Wikipedia. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2545–2555, Melbourne, Australia. Association for Computational Linguistics.

Larry M Bartels. 2006. Priming and persuasion in presidential campaigns. *Capturing campaign effects*, 1:78–114.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in NLP. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 995–1005, Jeju Island, Korea. Association for Computational Linguistics.

Amparo Elizabeth Cano-Basave and Yulan He. 2016. A study of the impact of persuasive argumentation in political debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1405–1413, San Diego, California. Association for Computational Linguistics.

Winston Carlile, Nishant Gurrapadi, Zixuan Ke, and Vincent Ng. 2018. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 621–631, Melbourne, Australia. Association for Computational Linguistics.

Zheqian Chen, Rongqin Yang, Zhou Zhao, Deng Cai, and Xiaofei He. 2018. Dialogue act recognition via crf-attentive structured network. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 225–234.

John W Cooley. 1993. A classical approach to mediation-part i: Classical rhetoric and the art of persuasion in mediation. *U. Dayton L. Rev.*, 19:83.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina 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, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. 2018. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1383–1392, Melbourne, Australia. Association for Computational Linguistics.

Esin Durmus and Claire Cardie. 2018. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1035–1045, New Orleans, Louisiana. Association for Computational Linguistics.

Ritam Dutt, Rishabh Joshi, and Carolyn Penstein Rose. 2020. Keeping up appearances: Computational modeling of face acts in persuasion oriented discussions. *arXiv preprint arXiv:2009.10815*.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Marieke L Fransen, Edith G Smit, and Peeter WJ Verlegh. 2015a. Strategies and motives for resistance to persuasion: an integrative framework. *Frontiers in psychology*, 6:1201.

Marieke L Fransen, Peeter WJ Verlegh, Amna Kirmani, and Edith G Smit. 2015b. A typology of consumer strategies for resisting advertising, and a review of mechanisms for countering them. *International Journal of Advertising*, 34(1):6–16.

Boris Galitsky, Dmitry Ilvovsky, and Dina Pisarevskaya. 2018. Argumentation in text: Discourse structure matters. *CICLing 2018*.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.

Teresa Heath, Robert Cluley, and Lisa O'Malley. 2017. Beating, ditching and hiding: consumers' everyday resistance to marketing. *Journal of Marketing Management*, 33(15-16):1281–1303.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R. Lyu. 2019. HiGRU: Hierarchical gated recurrent units for utterance-level emotion recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 397–406, Minneapolis, Minnesota. Association for Computational Linguistics.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2021. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In *International Conference on Learning Representations*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751.

Silvia Knobloch-Westerwick and Jingbo Meng. 2009. Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, 36(3):426–448.

Gaurav Kumar, Rishabh Joshi, Jaspreet Singh, and Promod Yenigalla. 2020. AMUSED: A multi-stream vector representation method for use in natural dialogue. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 750–758, Marseille, France. European Language Resources Association.

Liane Longpre, Esin Durmus, and Claire Cardie. 2019. Persuasion of the undecided: Language vs. the listener. In *Proceedings of the 6th Workshop on Argument Mining*, pages 167–176, Florence, Italy. Association for Computational Linguistics.

Andrew Luttrell and Vanessa Sawicki. 2020. Attitude strength: Distinguishing predictors versus defining features. *Social and Personality Psychology Compass*, 14(8):e12555.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825.

Vitobha Munigala, Abhijit Mishra, Srikanth G Tamilselvam, Shreya Khare, Riddhiman Dasgupta, and Anush Sankaran. 2018. Persuaide! an adaptive persuasive text generation system for fashion domain. In *Companion Proceedings of the The Web Conference 2018*, pages 335–342.

Daniel J O'Keefe. 2002. Guilt as a mechanism of persuasion. *The persuasion handbook: Developments in theory and practice*, pages 329–344.

Vipul Raheja and Joel Tetreault. 2019. Dialogue Act Classification with Context-Aware Self-Attention. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3727–3733, Minneapolis, Minnesota. Association for Computational Linguistics.

Victor Tejedor San José. 2019. The role of humor and threat on predicting resistance and persuasion.

Steven J Sherman and Larry Gorkin. 1980. Attitude bolstering when behavior is inconsistent with central attitudes. *Journal of Experimental Social Psychology*, 16(4):388–403.

Paul Surgi Speck and Michael T Elliott. 1997. Predictors of advertising avoidance in print and broadcast media. *Journal of Advertising*, 26(3):61–76.

Zakary L Tormala. 2008. A new framework for resistance to persuasion: The resistance appraisals hypothesis. *Attitudes and attitude change*, pages 213–234.

Anita L Vangelisti, John A Daly, and Janine Rae Rudnick. 1991. Making people feel guilty in conversations: Techniques and correlates. *Human Communication Research*, 18(1):3–39.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.

Wendy Wood and Carl A Kallgren. 1988. Communicator attributes and persuasion: Recipients' access to attitude-relevant information in memory. *Personality and Social Psychology Bulletin*, 14(1):172–182.

Peter Wright. 1975. Factors affecting cognitive resistance to advertising. *journal of Consumer Research*, 2(1):1–9.

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy. 2019. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3620–3630, Minneapolis, Minnesota. Association for Computational Linguistics.

Jichuan Zeng, Jing Li, Yulan He, Cuiyun Gao, Michael Lyu, and Irwin King. 2020. What changed your mind: The roles of dynamic topics and discourse in argumentation process. In *Proceedings of The Web Conference 2020*, WWW '20, page 1502–1513, New York, NY, USA. Association for Computing Machinery.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia. Association for Computational Linguistics.

Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019. A dynamic strategy coach for effective negotiation. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378, Stockholm, Sweden. Association for Computational Linguistics.

Julia Zuwerink Jacks and Kimberly A Cameron. 2003. Strategies for resisting persuasion. *Basic and applied social psychology*, 25(2):145–161.

Appendix

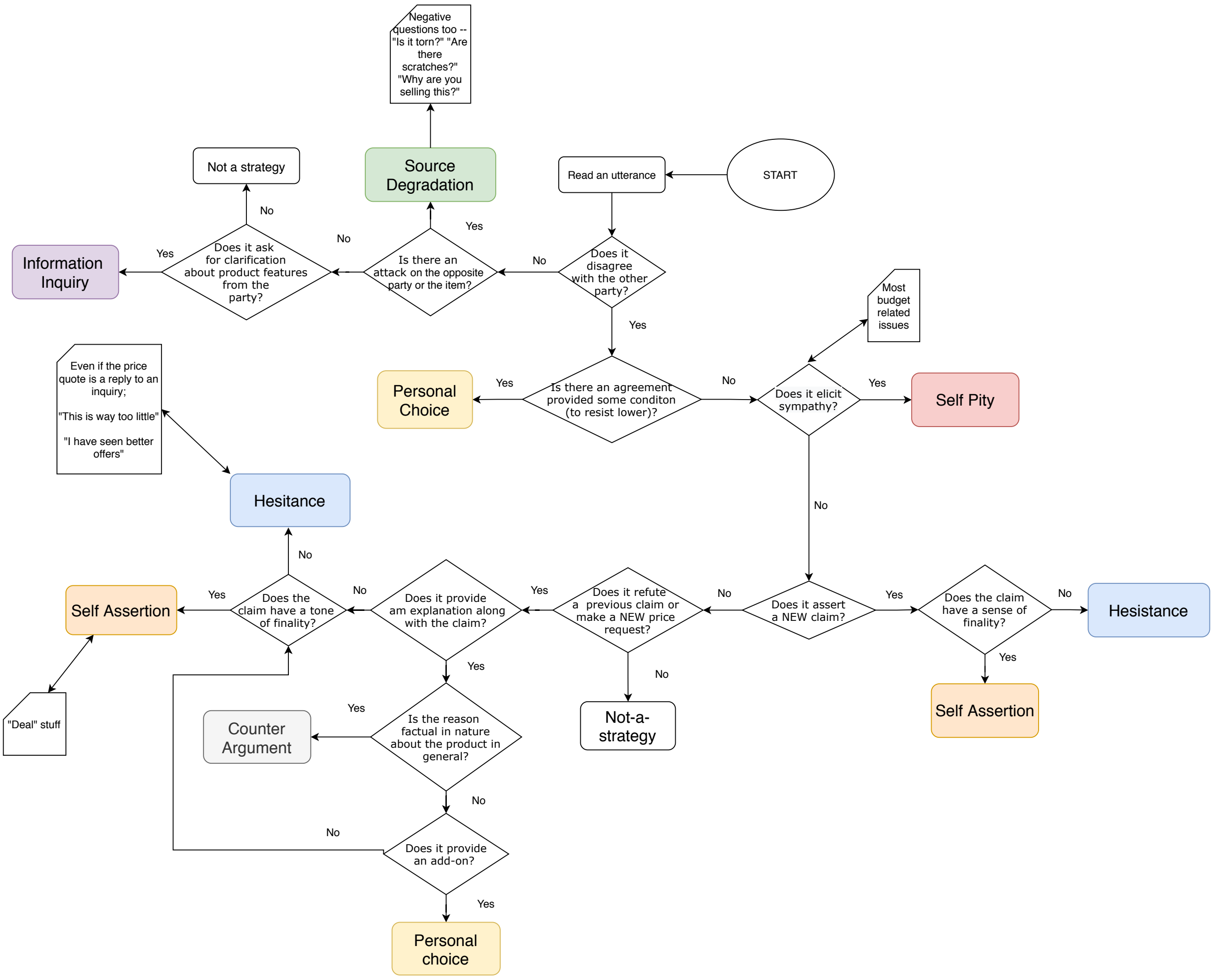

Figure 3: Flowchart for annotating CB