Automatically Labeled Data Generation for Classification of Reputation Defence Strategies

Nona Naderi and Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, ON, M5S 3G4, Canada
{nona,gh}@cs.toronto.edu

Abstract

Reputation defence is a form of persuasive tactic that is used in various social settings, especially in political situations. Detection of reputation defence strategies is a novel task that could help in argument reasoning. Here, we propose an approach to automatically label training data for reputation defence strategies. We experimented with over 14,000 pairs of questions and answers from the Canadian Parliament, and automatically created a corpus of questions and answers annotated with four reputation strategies. We further assessed the quality of the automatically labeled data.

1. Introduction

Maintaining good reputation is important in almost all social settings. Losing one's reputation can affect competitiveness, trust, position, and relations. Individuals, businesses, and institutions try to manage reputation threat or the danger of losing their reputation by using various persuasive defence strategies (Benoit, 1995). A recent prevailing example of reputation threat and defence is various sexual assault allegations and the use of strategies, such as *denial,* i.e., denying the situation, and *mortification,* i.e., the admission of guilt and apologizing and asking for forgiveness, in response to these allegations.

Maintaining good reputation is particularly important to politicians, as often the most acceptable political images are voted for and chosen by the electorate. Politicians who choose policies that impact citizens are more concerned with their reputations because they are held responsible for their actions by both citizens and other political parties.

One example of a reputation defence strategy is the expression of *mortification* in a statement that was issued by the U.S. Secretary of Health regarding the expenses of his travel on private planes:

"I regret the concerns this has raised regarding the use of taxpayer dollars. All of my political career I've fought for the taxpayers. It is clear to me that in this case, I was not sensitive enough to my concern for the taxpayer."[1]

While reputation defence strategies and their effectiveness have been extensively studied (Coombs and Holladay, 2008; Sheldon and Sallot, 2008; Burns and Bruner, 2000; Sheldon and Sallot, 2008; Lyon and Cameron, 2004), most of these studies are qualitative in nature. One exception is that of Naderi and Hirst (2017), who proposed a computational approach to identify reputation defence strategies from parliamentary debates. Here, we propose two semi-supervised approaches for identifying persuasive reputation defence strategies. One approach uses the observed word pairs from both reputation threat and reputation defence, and the other uses pattern-based representations of reputation defence.

We evaluated a subset of the automatically labeled data against crowd-sourced annotations. We further assessed the impact of the extended dataset in a multi-class classification task. We found that the approach based on the observed word pairs yields higher-quality labels for the reputation defence strategies.

2. Related Work

Ethos i.e., one's credibility has been considered as one of the important means of persuasion in Aristotle's rhetoric (Aristotle, 2007). In danger of losing credibility, one may prepare apologia that is a self-defence speech in response to the criticism or attack. According to Downey (1993), apologia has taken various functions and styles over time, for example, early contemporary apologia resembled classical apologia and used causal reasoning and detailed evidence; however, after 1960, apologia has been altering into "misleading narratives and dishonest apologies", replete with discrepancies. Similar to Downey's study, most previous work on persuasive reputation defence strategies focused on a few case studies (Brinson and Benoit, 1999; Benoit and Henson, 2009; Zhang and Benoit, 2009; Harlow et al., 2011) with the exception of one study. Naderi and Hirst (2017) created a corpus of parliamentary question and answers annotated with four reputation strategies and proposed a feature-based approach to detect these strategies (see Table 1). Parliamentary question periods provide a rich dataset to study various crises and the face-saving strategies that are used to manage these crises. Parliamentary question periods have previously been studied for analysing rhetorical aspects of questions (Zhang et al., 2017), interruption behaviour (Whyte, 2014), determining ideologies using party-membership (Hirst et al., 2014), and measuring emotions (Rheault et al., 2016).

While the task of automatic detection of reputation defence strategies is closely related to argumentation mining tasks (Stab and Gurevych, 2014; Nguyen and Litman, 2016; Biran and Rambow, 2011; Wang and Cardie, 2014; Peldszus, 2014), it differs in that it focuses on relations between arguments of reputation threat (questions)

[1] Health secretary Tom Price apologizes for taking private flights for work, *The Guardian,* 2017-09-28

Reputation defence strategies
Denial: 1. The government denies that the situation in question occurred. 2. The government denies causing the situation in question.
Excuse (evading responsibility): 1. The situation in question occurred in response to some other situations. 2. The situation in question occurred because of lack of information or control over important factors. 3. Some accidents caused the situation. 4. The motives or intentions of the government were good.
Justification (reducing offensiveness): 1. The government tries to increase positive feeling towards it (for example by mentioning positive actions the government performed in the past). 2. The government tries to convince the audience that the situation is not as bad they say. 3. The government tries to distinguish the situation in question from similar but less desirable situations. 4. The government tries to place the situation in a different or broader context. 5. The government attacks the opposition or questions their credibility. 6. The government offers compensation for the situation.
Concession (corrective actions): 1. The government promises to restore the situation to what it was before. 2. The government promises to make changes (for example to prevent the recurrence of the situation).
None of these strategies

Table 1: Conditions for each reputation defence strategy. This table is taken from the study by Naderi and Hirst (2017).

and reputation defence (answers). Previous studies on argumentation have shown that manually annotating argument-related information is difficult and results in moderate agreement (Habernal et al., 2014; Wachsmuth et al., 2017; Naderi and Hirst, 2017). Here, we aim to automatically create a large corpus of reputation defence strategies. We propose two approaches and examine the quality of the extracted data using these approaches.

3. Method

For our analysis, we used a dataset described by Naderi and Hirst (2017). This dataset consists of 493 pairs of questions and answers from Oral Question period from Canadian parliamentary proceedings, manually annotated with four reputation defence strategies (170 pairs of questions and answers are annotated as denial, 36 pairs as excuse, 173 pairs as justification, 95 pairs as concession, and 19 as none of these strategies). Here, we removed 19 pairs that were annotated as being *none* of these strategies, and focused on the remaining pairs. We refer to this corpus as the reputation defence strategy dataset throughout the paper. Given these manually labeled examples, we extracted a set of features to assign scores to unlabeled pairs of questions and answers and automatically expanded the training set.

3.1 Preprocessing of data

Here, we used the Lipad[2] (Linked PArliamentary Data) dataset (Beelen et al., 2017). This dataset consists of Canadian Hansards since 1901. We extracted 14,134 pairs of questions and answers from Oral Question period (1994–2014) as our unlabeled data. Since the questions asked by the government backbenchers are generally friendly and intended for clarification, we only focused on the questions asked by the opposition members and their respective answers by the government ministers. An example of a reputation defence is presented in Table 2. Furthermore, we extracted only the first question and answer pairs of each topic of discussion, because the remaining pairs require the context. We made sure that the pairs of questions and an-

Q. Mr. Speaker, we now know that the Prime Minister announced a $600,000 grant in his riding months before the project had been approved, and coincidentally just weeks before the federal election. Since only the Prime Minister knows when an election will be called, it is clearly and simply a case of announcing pre-election goodies. The Prime Minister would have us believe the grant was awarded after careful review, but program officer Lionel Bergeron thought differently when he said in a memo "This project has been announced by the Prime Minister. Its approval is urgent". How could the Prime Minister deny that he was just trying to influence voters in his riding by getting this grant before it went through the proper circle?
A. Mr. Speaker, this project had been discussed for years in Shawinigan. It is the kind of project that is badly needed in a district where unemployment is very high in the Saint-Maurice riding. Everyone had been talking about it. Everyone supported the project, including the hon. member for Saint-Maurice who has done his job as the local member for Saint-Maurice. We are very pleased that the project has worked and has indeed created the jobs that it was supposed to bring to the region.

Table 2: An example of reputation defence strategy; 1999-05-25, Chuck Strahl (Q) and Pierre S. Pettigrew (A).

[2] https://www.lipad.ca

swers from the reputation defence dataset were not included in our unlabeled dataset. We extracted two sets of features to assign scores to unlabeled question and answer pairs: (1) observed word pairs, (2) surface patterns. We will discuss these features in the following sections.

3.2 Pairs of words

Word pairs from a pair of arguments have been shown to be informative features in identifying implicit discourse relations between the two arguments (Marcu and Echihabi, 2002; Pitler et al., 2009; Biran and McKeown, 2013).
Additionally, Naderi and Hirst (2017) have shown that discourse relations between the question and answer sentences can help in capturing the relations between reputation threat and defence instances, and they can be informative features for the detection of reputation defence strategies. Therefore, we considered all the possible word pairs extracted from the cross-product of the question and answer. To represent the relevance of each word pair to each reputation defence strategy, we computed a correlation score using our seed examples. A score is assigned to each question and answer based on simple occurrences:

$$\left(\frac{\text{Count unique word pairs of Label}_i}{\text{Count total unique word pairs}}\right)$$

The raw score was then normalized by dividing by the sum of raw scores of all four strategies.

3.3 Pattern extraction

For extracting the surface patterns, we took an approach similar to that of Tsur et al., (2010). Using the extracted unlabeled question and answer pairs, we divided the words into frequent and infrequent words (IFW) according to their relative frequency in the unlabeled corpus and a specified threshold. This threshold was set to 1000 per million. The length of patterns was set to be 5 to 7 words with only 3 to 5 slots for infrequent words. Multiple patterns were extracted from each reputation defence answer. We then computed a score for each question and answer pair according to the exact matches of the patterns of each reputation defence strategy. For example, from the *denial* answer *Mr. Speaker, at no time have we interfered with the operations of Air Canada, and I stand by my answer of yesterday*, the following example patterns were extracted:

- *at no time have we* IFW *with*
- *no time have we* IFW *with the*
- *have we* IFW *with the*
- *i* IFW *by my* IFW *of yesterday*

Each question and answer pair was first assigned a raw score for each strategy, and then the score was normalized by the sum of all strategy scores (similar to the approach in Section 3.2):

$$\frac{\sum_k \text{Length}(\text{pattern}_k) \times \text{Count}(\text{pattern}_k)}{\sum_i \text{Score of Label}_i}$$

Q. Mr. Speaker, my question is for the Minister of Human Resources Development. It concerns the government's plans for the end to the TAGS program. How could the minister expect Canadians to take him seriously when he says that the government is working on plans to help out the affected communities after TAGS is finished and we know he is telling the RCMP and his own officials they should get ready for the fact that they will be doing nothing? The minister now has a copy of the leaked document before him. Will he explain why the government is making plans for a social disaster in fishing communities instead of preventing the end of assistance for fishing communities and the people in those areas?
A. Mr. Speaker, I have never asked the RCMP to do the sorts of things he said in his question. I understand that some of our officials need some training to be able to cope with confrontational situations and to handle more difficult situations on an individual basis. It has happened not only in relation to TAGS but across Canada. This is the way it works. Our government is doing the right thing by conducting a review of the post-TAGS situation. We are not particularly worried because we trust Canadians and we know Canadians behave properly all the time.

Table 3: An example of the *denial* strategy used together with the *justification* strategy; 1997-11-21, Peter Stoffer (Q) and Pierre S. Pettigrew (A).

Score of Label_i is a raw score of strategy i.
The extracted word pairs that were assigned highest scores based on the sets of features, patterns, or observed pairs of words were considered as candidates to be added to the training set.

4. Evaluation

In order to be able to examine the quality of the extracted candidates, we used a five-fold cross-validation approach for the extension and evaluation of the data. In each fold, we used 94 instances of the reputation defence dataset (Naderi and Hirst, 2017) for test, and the remaining for data extension (extracting patterns and observed word pairs from question and answer pairs) and classification task. We extended the training data once with only the observed word pairs, and once with only the pattern features. In each fold, the size of the training set varies according to the assigned scores. Since each answer can express multiple reputation strategies (see the example in Table 3) or none, we used a threshold value to decide whether to add the candidate pair to the training set or not. We examined various threshold values for each approach.
The quality of the extracted pairs was evaluated in two ways: (1) comparison with manual annotation, and (2) the contribution of the added training data to the classification of reputation strategies.

4.1 Inter-annotator agreement

To examine whether the assigned labels are of high quality, we conducted a study with 180 random question and answer pairs on the CrowdFlower platform. The ques-

(a) Does the answer express Concession?	(b) Does the answer express Justification?
Q. Mr. Speaker, my question is for the Minister of Labour. Former workers at Singer are arguing that the federal government did not fulfill its contract obligations toward them because it gave the company, instead of them, the Government Annuities Account surplus, that is a part of their pension funds that it was responsible for administering. Does the Minister of Labour not agree that the contract binding the parties between 1946 and 1957 is abundantly clear and that the federal government had an obligation to pay the surplus out to the workers and not to Singer? **A.** Mr. Speaker, all the federal regulations have been applied in this matter.	**Q.** Mr. Speaker, if we understand this correctly, 72% of Canada's refugee claimants have entered Canada from the United States of America, which means that 28% of refugees obviously come from refugee camps. Is the minister telling us that we are only accepting 28% of legitimate refugees to this country who actually deserve to be raised to higher levels? **A.** Mr. Speaker, the member is telling us that legitimate refugees are only people who we picked up, that everyone crossing our borders or arriving at our airports are not legitimate. He should be ashamed of himself.

Table 4: (a) Disagreement among six annotators, two of whom annotated it as *concession* and three as not *concession;* 1995-06-01, Claude Bachand (Q) and Lucienne Robillard (A). (b) Three of the annotators confirmed the answer as *justification* strategy and two as not *justification;* 2002-04-30, Rahim Jaffer (Q) and Denis Coderre (A).

All crowdsourced annotations

(a) Observed word pairs			
t > .33	**t > .32**	**t > .31**	**t > .30**
.60	.71	.73	.70

(b) Extracted patterns			
t > .90	**t > .80**	**t > .70**	–
.41	.43	.43	–

Crowdsourced annotations with confidence > 80%

(c) Observed word pairs			
t > .33	**t > .32**	**t > .31**	**t > .30**
.80	**.85**	.77	.76

(d) Extracted patterns			
t > .90	**t > .80**	**t > .70**	–
.41	.39	.38	–

Table 5: (a) Evaluation of automatically assigned strategies using observed word pairs against all crowd annotations; (b) Evaluation of automatically assigned strategies using extracted patterns against all crowd annotations; (c) Evaluation of automatically assigned strategies using observed word pairs against crowd annotations with confidence > 80%; (d) Evaluation of automatically assigned strategies using extracted patterns against crowd annotations with confidence > 80%. t is the threshold used for accepting the candidate labels.

tion and answer pairs were sampled from a pool of pairs that were assigned a reputation strategy label using the two approaches that were described earlier (see Sections 3.2 and 3.3).

Contributors were shown a question and answer pair with the assigned reputation defence strategy, as well as the description and conditions of the assigned strategy from Table 1. The contributors were then asked whether the assigned strategy was correct or not. We asked for at least five annotations per pair from the English-speaking countries. The contributors were presented with one test pair of question and answer and three other pairs on each page, and had to maintain 80% accuracy throughout the job. In total, the task included 66 *denial,* 5 *excuse,* 79 *justification,* and 30 *concession* questions. 81 of 180 were agreed by all 5 annotators. Only 59 answers were annotated with a confidence score below 80%. The confidence score is the agreement of the five annotators weighted by the annotators' trust scores.[3] Trust scores are determined by the annotators' accuracy on the test questions they have seen. Table 4 shows two examples of disagreement by the annotators. Most of the answers that caused disagreement among annotators evaded providing a response to the given question.

Table 5 shows what percentage of the automatically assigned strategies using word pairs and pattern acquisition approaches were correct compared to the crowdsourced annotations. We once considered all the crowdsourced data. We further removed the crowdsourced annotations with the confidence scores lower than 80%, and assessed the quality of the automatically assigned labels against higher-quality crowdsourced annotations. When compared with the crowdsourced annotations with a confidence score of at least 80%, the labels that were extracted using the observed word pairs approach with the threshold $t > .32$ shows the highest agreement. The automatically assigned labels using pattern acquisition approach show low agreement with the crowdsourced annotations.

4.2 Five-fold cross-validation

We further evaluated the quality of the data by assessing its contribution to the classification task. As mentioned earlier, we performed a five-fold cross-validation using the reputation defence dataset. The test set always came from the reputation defence dataset. We performed a multi-class classification using a class-weighted Support Vector Ma-

[3] https://success.crowdflower.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score

	Original	**t > .33**	**t > .32**	**t > .31**	**t > .30**
Train	379	512	1238	3797	8495
		BOW			
F_1	51.32	54.65	55.39	52.61	55.28
Accuracy	53.35	56.74	59.10	56.32	62.00
Denial	62.40	64.86	65.69	63.29	75.77
Excuse	13.60	17.00	13.64	13.64	3.64
Justification	55.60	62.42	66.39	63.50	67.14
Concession	36.40	32.00	25.00	14.32	11.02
		BOW+Negation+VerbNet+Similarity+Senti.+Disc.			
F_1	56.92	55.62	54.83	51.86	56.42
Accuracy	57.59	57.37	57.58	55.48	62.85
Denial	65.00	64.73	64.82	63.83	76.60
Excuse	18.00	17.00	17.27	17.00	6.60
Justification	59.80	62.30	64.75	63.05	67.50
Concession	48.00	37.74	24.30	13.01	10.80
		BOW+Negation+VerbNet			
F_1	53.22	54.77	56.01	53.05	55.29
Accuracy	54.22	56.11	58.84	56.74	62.01
Denial	63.60	64.73	65.60	63.45	75.95
Excuse	17.80	14.97	17.27	13.63	3.64
Justification	56.40	60.17	65.63	63.78	67.20
Concession	39.80	36.32	27.56	16.39	10.68

Table 6: Classification of reputation defence strategies using the extended training data with observed word pairs. The performance of classification of each strategy is reported in terms of average F_1. *t* is the threshold used for accepting the candidate labels.

	Original	**t > .90**	**t > .80**	**t > .70**
Train	379	453	486	573
		BOW		
F_1	51.32	48.52	47.63	47.51
Accuracy	53.35	49.99	49.15	48.94
Denial	62.40	56.94	54.73	57.54
Excuse	13.60	13.60	11.64	17.00
Justification	55.60	53.44	53.76	52.53
Concession	36.40	34.83	33.60	29.35
		BOW+Negation+VerbNet+Similarity+Senti.+Disc.		
F_1	56.92	49.00	49.10	49.53
Accuracy	57.59	50.84	50.62	51.26
Denial	65.00	56.60	56.01	56.61
Excuse	18.00	13.60	9.40	12.53
Justification	59.80	54.00	54.15	54.65
Concession	48.00	38.60	39.73	40.17
		BOW+Negation+VerbNet		
F_1	53.22	49.81	48.55	48.18
Accuracy	54.22	51.25	49.90	49.36
Denial	63.60	58.10	57.24	57.79
Excuse	17.80	18.10	18.10	27.42
Justification	56.40	54.32	53.30	51.66
Concession	39.80	36.94	34.58	30.89

Table 7: Classification of reputation defence strategies using the extended training data with patterns. The performance of classification of each strategy is reported in terms of average F_1. *t* is the threshold used for accepting the candidate labels.

chine model with a linear kernel[4] and the features proposed by Naderi and Hirst, including answer bag-of-words representations (weighted using *tf-idf*) of the answers, VerbNet verb classes, positive and negative sentiments, and negations in the answers, as well as discourse relations and similarity measure between the question and answer. We extracted the sentiments using OpinionFinder (Wilson et al., 2005) and discourse relations using End-to-End PDTB-Styled Discourse Parser (Lin et al., 2014). We further used the word2vec embeddings (Mikolov et al., 2013) for computing the similarity between the questions and answers.

Table 6 shows the results of the classification with the extended data using the observed word pairs approach. We used various threshold values (*t*) for accepting the candidates for the extension of the training data (train). Since in each fold the size of the extended data varies, we report the average size of the training sets of all folds. The baseline is the original dataset without any added data (the column specified as *original* in Table 6). The average F_1 measure of each reputation defence strategy is also presented. As shown in the table, by adding the automatically assigned labels to the training set, the performance of the classification of the *denial* and *justification* strategies improves; however, the data extension does not improve the classification of the *excuse* and *concession* strategies. Examining the extended data, we find that most of the added instances are *denial* and *justification* instances, and only a few pairs of questions and answers are annotated with the *excuse* and *concession* strategies. The reputation defence dataset consists of the total of only 36 *excuse* and 95 *concession* annotations; thus it is expected that the extended dataset includes very few of these strategies. Using the automatically added labels, the average F_1 measure of *denial* and *justification* reaches about 75% and 67%, respectively.

When we added the discourse relation and sentiment features, we did not observe any improvement in classification for the extended data. This can be due to having noise in the automatically assigned labels, and also the noisy nature of discourse relations and sentiment annotations.

Table 7 presents the results of the classification with the extended data using pattern acquisition approach. Extending the data using this approach does not result in a high-quality dataset and the performance of the classification drops very quickly. To improve the quality of the labels, we further examined whether removing the patterns that appreared in all the other strategies help. For example, for denial, we removed the patterns that appreared in non-denial examples. After removing the patterns that were shared between different strategies, we computed the scores introduced in Section3.3; however, we did not observe any improvements. Reputation defence strategies do not apply to all question and answer pairs (see the example in Table 8), and although we removed the few question and answer pairs annotated with *none* from the seed examples, we might be able to find these cases using a threshold value for accepting the candidate labels.

[4]LibSVM implementation (Pedregosa et al., 2011).

Q. Mr. Speaker, a week after the latest escalation in the conflict in Bosnia, when 370 peacekeepers, including 55 Canadians, were taken hostage by Serbian forces, there has been a flurry of statements and meetings which failed to produce any concrete results leading to the release of the hostages. This morning, the International Red Cross said that the Bosnian Serbs told them they would release the hostages unconditionally, either today or tomorrow. Could the Deputy Prime Minister confirm the statement by the Red Cross that the Bosnian Serbs will release the 370 peacekeepers who are being kept hostage sometime during the next few hours, although Bosnian Serb leader Radovan Karadzic said yesterday that no hostages could be released without guarantees that all air strikes would be suspended?
A. Mr. Speaker, we received communications mentioning that a few hostages might be released today, but at 11.13 a.m., we were unable to confirm whether that was the case.

Table 8: An example of an answer where *none* of the strategies apply; 1995-06-02, Gilles Duceppe (Q) and Sheila Copps (A).

5. Conclusion

We presented two approaches to automatically induce a corpus of reputation defence strategies. We considered pattern-based representation of reputation defence strategies and the observed pairs of words from the cross-product of questions and answers. We evaluated the generated data using the two proposed approaches against crowd annotation, and also assessed its contribution in the classification task. The observed word pairs approach resulted in a higher quality dataset. We found that the extended dataset using the observed word pairs contributes positively to the performance of the classifier, even though it contains noisy and weak labels.

Acknowledgements

This research is financially supported by the Natural Sciences and Engineering Research Council of Canada. We thank the anonymous reviewers for their suggestions.

Bibliographical References

Aristotle, Translated by Kennedy, G. (2007). *On Rhetoric: A Theory of Civic Discourse*. Oxford University Press.

Beelen, K., Thijm, T. A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., Lijbrink, S., Marx, M., Naderi, N., Rheault, L., Polyanovsky, R., and Whyte, T. (2017). Digitization of the Canadian parliamentary debates. *Canadian Journal of Political Science*, 50(3):849–864.

Benoit, W. L. and Henson, J. R. (2009). President Bush's image repair discourse on Hurricane Katrina. *Public Relations Review*, 35(1):40–46.

Benoit, W. L. (1995). *Accounts, Excuses, and Apologies: A Theory of Image Restoration Strategies*. State University of New York Press, Albany.

Biran, O. and McKeown, K. (2013). Aggregated word pair features for implicit discourse relation disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 69–73, Sofia, Bulgaria. Association for Computational Linguistics.

Biran, O. and Rambow, O. (2011). Identifying justifications in written dialogs. In *Proceedings of the 2011 IEEE Fifth International Conference on Semantic Computing*, pages 162–168, Washington, DC, USA. IEEE Computer Society.

Brinson, S. L. and Benoit, W. L. (1999). The tarnished star: Restoring Texaco's damaged public image. *Management Communication Quarterly*, 12(4):483–510.

Burns, J. P. and Bruner, M. S. (2000). Revisiting the theory of image restoration strategies. *Communication Quarterly*, 48(1):27–39.

Coombs, W. T. and Holladay, S. J. (2008). Comparing apology to equivalent crisis response strategies: Clarifying apology's role and value in crisis communication. *Public Relations Review*, 34(3):252–257.

Downey, S. D. (1993). The evolution of the rhetorical genre of apologia. *Western Journal of Communication*, 57(1):42–64.

Habernal, I., Eckle-Kohler, J., and Gurevych, I. (2014). Argumentation mining on the web from information seeking perspective. In Elena Cabrio, et al., editors, *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, pages 26–39. CEUR-WS.

Harlow, W. F., Brantley, B. C., and Harlow, R. M. (2011). BP initial image repair strategies after the Deepwater Horizon spill. *Public Relations Review*, 37(1):80–83.

Hirst, G., Riabinin, Y., Graham, J., Boizot-Roche, M., and Morris, C., (2014). *From Text to Political Positions: Text analysis across disciplines*, chapter Text to ideology or text to party status?, pages 61–79. John Benjamins Publishing Company, Amsterdam.

Lin, Z., Ng, H. T., and Kan, M.-Y. (2014). A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20(2):151–184.

Lyon, L. and Cameron, G. T. (2004). A relational approach examining the interplay of prior reputation and immediate response to a crisis. *Journal of Public Relations Research*, 16(3):213–241.

Marcu, D. and Echihabi, A. (2002). An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 368–375, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Naderi, N. and Hirst, G. (2017). Recognizing reputation defence strategies in critical political exchanges. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 527–535, Varna, Bulgaria.

Nguyen, H. and Litman, D. (2016). Context-aware argumentative relation mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers), pages 1127–1137, Berlin, Germany. Association for Computational Linguistics.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peldszus, A. (2014). Towards segment-based recognition of argumentation structure in short texts. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics.

Pitler, E., Louis, A., and Nenkova, A. (2009). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 683–691, Suntec, Singapore. Association for Computational Linguistics.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. *PloS one*, 11(12):e0168843.

Sheldon, C. A. and Sallot, L. M. (2008). Image repair in politics: Testing effects of communication strategy and performance history in a faux pas. *Journal of Public Relations Research*, 21(1):25–50.

Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 46–56, Doha, Qatar. Association for Computational Linguistics.

Tsur, O., Davidov, D., and Rappoport, A. (2010). ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In William W. Cohen et al., editors, *International AAAI Conference on Weblogs and Social Media*. The AAAI Press.

Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., Hirst, G., and Stein, B. (2017). Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187, Valencia, Spain. Association for Computational Linguistics.

Wang, L. and Cardie, C. (2014). Improving agreement and disagreement identification in online discussions with a socially-tuned sentiment lexicon. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 97–106, Baltimore, Maryland. Association for Computational Linguistics.

Whyte, T. (2014). Some honourable members: A quantitative analysis of parliamentary decorum in canada and the united kingdom. In *23rd World Congress of Political Science*, Montreal, Canada.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.

Zhang, E. and Benoit, W. L. (2009). Former Minister Zhang's discourse on SARS: Government's image restoration or destruction? *Public relations review*, 35(3):240–246.

Zhang, J., Spirling, A., and Danescu-Niculescu-Mizil, C. (2017). Asking too much? The rhetorical role of questions in political discourse. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1558–1572, Copenhagen, Denmark. Association for Computational Linguistics.