Contextual Cues for Agreement and Rejection in Spoken Dialogue

Julian Schlöder

Abstract

To understand and analyze a dialogue, it is important to keep track of the common ground the speakers build up. Therefore it is important to be able to discern between *acceptance* and *rejection* of propositions. Previous work was often concerned with written dialogue, but spoken dialogue differs from written dialogue in that it contains a wealth of additional information, *e.g.*, disfluencies or meta-communicative acts like backchannels. That turn taking is a collaborative act also gives additional importance to the length of statements. We theoretically motivate how this additional information can be used to discern between acceptance and rejection. We then verify our theoretical considerations with data from the Switchboard corpus and by using them to derive features for a classifier; our classifier reaches a 73% recall of rejections, an improvement of 14pp over a simple baseline.

1 Introduction

A central concept in dialogue analysis is the notion of *common ground* which contains proposals the speakers A and B have agreed on (Fernandez, 2013). To keep track of the contents of the common ground, it is necessary to identify which *proposals* made by one speaker are accepted (or rejected) by the other one; for a rejected proposal is not added to the common ground (Stalnaker, 2013).

The surface form of an acceptance or rejection is *not* always explicit and must be inferred (Walker, 1996). However, a rejection is frequently considered a *dispreferred response* (McTear, 2004) which has a recognizable impact on the continuation of the dialogue, *e.g.*, a dispreferred response is more likely to be initiated with a word like *well* (Byron and Heeman, 1997).

Spoken dialogue contains a lot of meta-information; the speakers are able to exchange online feedback whereas in written dialogue both dialogue partners type up their respective responses on their own. We expect that the differences between agreement and rejection will have an impact visible in these mechanisms, *e.g.*, backchannels and turn taking. Furthermore, speech disfluencies may also give additional insight.

We intend to show that these *structural* and *contextual* properties of the surrounding turns such as turn and utterance length, backchannels and disfluencies are connected to acceptance and rejection and that these cues can help in related classification tasks.

In the long run, reliable identification of acceptance and rejection may help in summarization tasks (Galley et al., 2004). For example, if one is interested in summarizing the outcome of an discussion or debate, the points where the speakers have reached consensus is of interest.

We first give an overview of the related work in section 2. In section 3 we we discuss the theoretical motivations for structural and contextual cues. We check our hypothesis by gathering appropriate data from the Switchboard corpus and discuss this data in section 4. We then select features for a Bayesian Classifier based on our theoretical and data-based observations and evaluate the results in section 5. Finally, section 6 contains a reflection on our results, thoughts on their reliability and their shortcomings and some possibilities for further improvement and research.

2 Related Work

Previous approaches in acceptance/rejection classification have focussed on utterance- or turn-internal features, *e.g.*, using to cue words (Misra and Walker, 2013). In this paper we search for useful features in the turns surrounding the acceptance/rejection. Furthermore, we concentrate mainly on structural observations which are largely independent of the wording of the acceptance or rejection, using cue words merely as our baseline.

Some work has been done in this area using the ICSI Meeting Corpus (Janin et al., 2003), a partly annotated corpus of spoken, multi-person discussions. Taking contextual features into account improved the accuracy (in the task of recovering acceptance/rejection/backchannel/other type utterances) from 80% (Hillard et al., 2003) to 86.9%.

We apply related (though simpler) features in a similar classification task (discerning acceptance from rejection) in the Switchboard corpus. These papers are also concerned with speaker identification in a multi-person discussion; a problem that does not arise in the Switchboard corpus.

There are much more rich written corpora; for discussion and debate in particular there is the Internet Argument Corpus IAC (Walker et al., 2012). Context-independent features have reached a accuracy of 66% in discerning agreement from rejection (Misra and Walker, 2013).

The task is close to what we consider: discerning acceptance from rejection from a set known to consist of only these two. However, the contextual and structural features we consider are largely unique to spoken dialogue and not applicable to the IAC.

Going beyond the features added by spoken dialogue, it is also possible to use nonverbal, audiovisual cues for a task like that (Bousmalis et al., 2009).

In a wider context, there are results about acceptance and rejection from viewpoints closer related to inference (Walker, 1996) and reasoning (Biran and Rambow, 2011). We did not take these observations into account, but we will give some pointers as to how they might improve our data.

3 Characteristics of Acceptance and Rejection

3.1 Cue Words

The first and most intuitive approach in classifying acceptance and rejection is the search for appropriate *cue words*, the most overt ones being "yes" (indicating acceptance) and "no" (indicating rejection). Similar phrases are "yeah", "right", "agree" or "don't", "not", "can't". But there are more subtle cue words. Consider the following example:

```
aa A: I was never into those
movies, either./
```

While the utterance "I was never into those movies" could be just as well a rejection, the cue word "either" identifies it unambiguously as an acceptance. Similarly, the word "actually" is a clear-cut cue for rejection. A rather special case is the hedge word "well", typically initiating a rejection (Byron and Heeman, 1997). Depending on the corpus, there can be a rather large number of useful unigram cue words and ngram cue phrases (Misra and Walker, 2013).

Unfortunately, cue words alone are no surefire way to determine whether or not an utterance is accepting or rejecting. Intuitively, one might assume that "yes" and "no" are words definitely implying acceptance or rejection respectively. However, consider for example the following two excerpts from the Switchboard corpus:

```
sd A: {C But, } yet, Parkland
is not free, either.  /
aa B: No, /
sd B: it's not free, /

sv B: {C or } they wouldn't be
able to own a house.  /
ar A: Yes, /
ar A: they would.  /
```

Here, "No" is an accepting utterance and "Yes" a rejecting one. Sometimes this also leads to the cue words outright contradicting each other, *e.g.*,

```
sv A: I don't think anybody pays
too little.  /
aa B: No, /
aa B: I would tend to agree very
highly.  /
```

Apparently, negative statements can be rejected by "yes" and accepted by "no". In the latter case, this seems to be form of logical *double negation*. In the former case this seems to be a form of *contraposition*: Just as a positive statement is rejected by a negative phrase, a negative phrase can be rejected by a positive phrase. Double negation and contraposition are *contextual cues* as to whether a utterance is accepting or rejecting.

The primary observation is that it is of central importance to take the context of an utterance into account if one wants to determine whether or not the utterance is accepting or rejecting. Nevertheless, in written debates, unigram cue words have achieved an accuracy of 60% (Misra and Walker, 2013). We will now make the case for taking more, and more subtle, contextual cues into account.

3.2 Additional and Contextual Cues

Speakers actively collaborate on extending their common ground (Fernandez, 2013). Thus while in agreement, the listener will possibly let the speaker continue and signal his continued agreement with backchannels, while he will take the turn to reject once a disagreement occurs.

There would be nothing to gain if the speaker would elaborate on something the listener will reject, so it is in the interest of both that rejections are voiced as early as possible. Therefore we expect that turns preceding agreement are longer with more backchannels from the listener. Furthermore, the dialogue partners should show different reactions to agreement or disagreement that will be visible in, *e.g.*, speech disfluencies or turn change behaviour.

Similarly, in spoken dialogue the speakers utilize meta-communicative feedback acts to signal successful grounding (Fernandez, 2013). It may be expected that a speaker that is about to reject a proposal would give lesser (or no) such feedback acts, *i.e.*, we may see a reduction in backchannels before a rejection.

It has been observed that rejections are *dispreferred* responses, and as such require careful phrasing as not to seem impolite (McTear, 2004). This carefulness alone leads us to expect that a rejecting utterance will be longer than an accepting one. A notable additional observation is that rejections are frequently started with hedges such as “well” (Byron and Heeman, 1997), or “actually”, “rather” (Misra and Walker, 2013).

In addition, the rejecting speaker might want to justify giving a dispreferred response; this also leads us to expect rejecting utterances (and the turns they occur in) to be comparatively longer. As a dispreferred response is also *unexpected* (Bousfield, 2008) a rejection may also cause an increase in disfluencies, as the speaker needs time to get accustomed to an unexpected turn in the dialogue.

4 Acceptance and Rejection in the Switchboard Corpus

4.1 Overview

The Switchboard corpus is a collection of around 2400 recorded and transcribed telephone conversations between two people (Godfrey and Holliman, 1997). The speakers are provided with a topic and then converse freely. The corpus has been annotated with the SWBD-DAMSL labels.

We do not take all *aa* (accept) and *ar* (reject) type dialogue acts from the Switchboard corpus into consideration. Since we are interested in identifying what information will be added to the common ground, we select for 2 criteria:

i. The aa/ar utterance follows a turn in which a dialogue act of the type sd (statement-non-opinion), sv (statement-opinion), bf (summarize/reformulate) or ad (action-directive) was made.

ii. The aa/ar utterance is the first utterance in the speakers turn.

The second criterion discards roughly 50% of all candidate aa/ar type utterances; this is mostly due to speakers repeating such utterances, *e.g.*:

```
sd A: <Laughter> They live close
then.  /
aa B: Yeah, /
aa B: real close.  /
sd B: I go there, /
```

For each utterance that satisfies the criteria we gather data on three turns in the dialogue: The turn that utterance initiates (the *current turn*), the immediately preceding turn (the *preceding turn*) and the immediately following turn (the *next turn*). So if, *e.g.*, speaker A has made a statement φ and B rejected it, the preeeding turn is B's turn that ended with φ, the current turn is A's turn that started with the rejection, and the next turn is B's response to that.

We are most interested in length, backchannels and disfluencies. We collected data on length and disfluencies in the accepting/rejecting utterance, and data about length, disfluencies and backchannels in the surrounding turns. To correct for length and number of utterances, disfluencies were computed *per word* and backchannels were computed *per utterance*.

4.2 Evaluation

The data we extracted from the switchboard corpus is in tables 1 to 4. A few of the datapoints are particularly striking given our theoretical considerations.

	Absolute		Percentages	
	Acc.	Rej.	Acc.	Rej.
Total Utt.	6805	154	100%	100%
One-w. Utt.	4529	79	66.56%	51.3%
$\geq$ 4-w. Utt.	430	34	6.3%	22.1%
$\geq$ 7-w. Utt.	82	17	1.2%	11.0%
Avg. length	1.62	2.88	-	-
Disf. per Word	0.05	0.14	-	-

Table 1: Accept/Reject Utterance Features.

As expected, the rejecting utterances are longer (on average by one word) than the accepting ones, and increasing length is an increasingly good indicator of rejection, cf. table 1 in the ≥ 4 and ≥ 7 rows. Also in accordance with our predictions is the higher number of disfluencies per word in rejecting utterances. This figure is partly due to the fact that the Switchboard corpus treats an initial hedging "Well," as a disfluency; rejections frequently start with such a hedge, for example

```
ar A: {D Well } no, /
```
or
```
ar A: {D Well, } they did have
one woman dean.  /
```

	Accept	Reject
Avg. Utterances	3.40	3.86
Avg. Words	22.04	26.01
Avg. Backchannels per Utt.	0.07	0.08
Avg. Disfluencies per Word	0.07	0.09

Table 2: Current Turn Features.

The current turn is also comparatively longer (around 0.4 utterances and 4 words on average) in case of a rejection. This fits in our theoretical notion that a rejecting speaker uses his turn to make a counterpoint, whereas an accepting speaker would yield her turn more easily.

Interestingly, there is no significant difference in backchannels or disfluencies. Apparently, the rejecting speaker goes back to a normal amount of disfluencies quite quickly after the initial rejection.

	Accept	Reject
Avg. Utterances	3.91	3.43
Avg. Words	36.04	27.39
Avg. Backchannels per Utt.	0.14	0.09
Avg. Disfluencies per Word	0.08	0.08

Table 3: Preceding Turn Features.

We also find a difference in length in the preceding turn; in case of a rejection this turn is on average around 0.5 utterances and roughly 8.5 words shorter. Also, if the other speaker will make a rejection in his next turn, she will give fewer backchannels beforehand. This backs up the thought that the speaker about to reject a proposal will signal less understanding beforehand and take the turn earlier.

On the other hand, the speaker making a rejected proposal is apparently not aware that she is making a controversial or dispreferred statement: her speech is not more (or less) disfluent than when making an accepted proposal.

	Accept	Reject
Avg. Utterances	3.02	2.51
Avg. Words	24.85	17.08
Avg. Backchannels per Utt.	0.10	0.05
Avg. Disfluencies per Word	0.11	0.15

Table 4: Next Turn Features.

In the next turn we again find a difference in length. The next turn is on average around 0.5 utterances and 7.5 words shorter if following a rejection than when following an acceptance. There also seems to be an increase in disfluencies after an rejection and a decrease in backchannels. These results are neither in accordance not in contradiction to our theoretical considerations.

4.3 Data Quality

Unfortunately, from a statistical analysis standpoint, the data is not of a very high quality. Looking at table 1 we find an enormous majority of acceptances in comparison to the rejections. So while we might think that we have sufficient acceptances to smooth out outliers, this is certainly not the case for the rejection.

Accept / Reject	Avg	σ	Max	50^{th}
Avg. length	1.62	1.3	24	1
Disf. per W.	0.05	0.2	1.0	0.0
Prec. Avg. W.	36.04	38.2	462	23
Next Avg. W.	24.85	33.3	400	12
Avg. length	2.88	3.5	25	1
Disf. per W.	0.14	0.2	1.0	0.0
Prec. Avg. W.	27.39	29.1	151	18
Next Avg. W.	17.08	27.9	185	6

Table 5: Statistical Data. σ is the standard deviation and 50^{th} is the median.

In fact, every difference between acceptances and rejections observed in tables 1 to 4 is less then one standard deviation. We found the length-related categories for the current utterance and the preceding and following turns to be particularly striking, as well as the disfluencies per word in the current utterance. Table 5 shows statistical data on these datapoints.

The high standard deviation makes the data insignificant from a quantitative standpoint; the maximal datapoints show the existence of far-removed outliers. The noticeable difference between average and median is also evidence for severely skewed data.

However, the qualitative observations we made seem to be supported by the data nevertheless. The qualitative differences we observed in the averages are also visible in the medians and the maxima.

It may be possible to clean up the data for the acceptances (filter out outliers), but the set of rejections is too small for that. Furthermore, this data might indicate that there are different types of acceptances and rejections and it is uncalled for to gather data on all acceptances/rejections collectively. Again, a deeper investigation is hindered by the low number of rejections in the Switchboard corpus. There are, however, approaches that consider distinct types of acceptances/rejections (Walker, 1996).

5 Experiment

5.1 Design and Baseline

We take the set of acceptances and rejections as described in section 4.1 as our population and aim to discern between the two. The immense majority of acceptances over rejections poses some challenges: the trivial approach – just classifying every utterance in the set as an acceptance – produces results that seem ostensibly reasonable, *e.g.*, an accuracy of 97%.

It makes sense – also with a possible application in summarization tasks in mind – to consider the classification problem in terms of information retrieval: We want to see which proposals are committed to the common ground, so we need to find which ones have been rejected. We may also define the task as finding the acceptances, but due to their majority we will even with the trivial approach obtain precision and recall well above 95%. Precision and recall computed for the task of finding rejections are by far the most discerning values.

So we define our information retrieval task as retrieving the rejections from the set of acceptances and rejections. Our goal is to extract the relevant utterances (rejections) with as few false negatives (*i.e.*, rejections classified as acceptances) as possible. That is, we want to achieve high recall without losing too much precision.

Because of the limited data available, we used a Bernoulli distributed naive Bayesian classifier: `BernoulliNB` from the `scikit-learn` Python library (Pedregosa et al., 2011); the documentation of `scikit-learn` (scikit-learn developers, 2013) indicates that this classifier is well-suited for sparse data. Indeed, it produced the best result of all classifiers from `scikit-learn` we tested.

We tested each featureset by crossvalidating the classifier 10-fold on A, each time taking 90% of the data as a training set and the remaining 10% as test data. We compute the recall and precision (with respect to classification as rejection) as an average of the 10 runs.

For a naive baseline we select[1] cue words from section 3.1 and train the classifier on their appearance in the first two utterances of the *current turn* (if the current turn is more than one utterance long); we call this featureset CUE WORDS.

The classifier trained on this featureset would (on each training set) classify each utterance in the test set as an acceptance, thereby reaching recall of 0%. It is unsurprising that cue words have no effect; as observed in in section 3.1 some cue words can occur with both acceptance and rejection, so they are simply not very discerning features on their own.

[1] In no particular order, “yeah”, “right”, “yes”, “no”, “don’t”, “not”, “actually”, “either”, “correct”, “accept”, “agree”, “well”, “can’t”, “wouldn’t”.

So we amend the naive baseline according to the observations in section 3.1 with features checking double-negation and negation-affirmation. We define the set *never, no, not, *n't*[2] as negation-indicators and *yeah, right, yes* as affirmation-indicators. Note that negation/affirmation-indicators overlap but differ from rejection/acceptance-cue words: Cue words like "either", "agree" or "actually" are always indicative of agreement/rejection no matter the context.

We define the feature `double-negation` as the occurrence of an negation-indicator where the last utterance of the preceding turn contained a negation-indicator, and the feature `single-negation` as the occurrence of an negation-indicator where the last utterance of the preceding turn did not contain a negation-indicator. Analogously, define the features `double-affirmation` and `single-affirmation`. Call the featureset containing these four NEGATION.

The classifier trained on CUE WORDS+NEGATION yielded a precision of 52.68% with a recall of 59.35%.

5.2 Feature Selection

The most striking differences we isolated in section 4.2 were features of length, in particular the length of the utterance constituting the acceptance or rejection. Based on the data in table 1 we define the feature `geq4` as the accepting/rejecting utterance having 4 or more words, and the feature `geq7` as the accepting/rejecting utterance having 7 or more words. We call these features together the featureset LENGTH.

The lengths of the preceding and next turns were also significantly different. We adopted the medians we computed in section 4.3 as significant points. The featureset TURNS consists of the features `next_geq6`, `next_geq12`, `next_geq18` meaning that the next turn consists of 6/12/18 or more words, and the features `prec_geq13`, `prec_geq18`, `prec_geq23` meaning that the preceding turn consists of 12/18/23 or more words.

We were unable to extract any other useful features, despite the data in section 4.2 showing interesting effects related to backchannels and disfluencies. The most overt problem is that the classifier we used only accepts binary true/false features, whereas the "per utterance" or "per word" data we collected would be suited to define a feature describing a continuum or spectrum.

Also, disfluency-related features in the accepting/rejecting utterance showed little effect, because a large number of the disfluencies in rejections are hedges which are already covered by the cue words.

5.3 Results

Table 6 shows the precision and recall of rejection-type utterances using the various feature sets (and combinations thereof) we defined.

Features	Prec. (ar)	Rec. (ar)
Cue Words	-	0.0%
CW + Negation	52.68%	59.35%
CW + Ng. + Length	50.06%	65.05%
CW + Ng. + Turns	41.82%	70.78%
CW + Ng. + Ln. + Turns	43.71%	73.69%

Table 6: Experiment Data

Both of the featuresets TURNS and LENGTH show a noticeable difference to the baseline; in both cases the recall is much improved, and the precision is reduced. Using LENGTH brings a rather big improvement in recall (around 5.5pp) with an agreeable reduction in precision (around 2.5pp), whereas TURNS results in an improvement and a reduction of roughly 11pp each.

We would have hoped that LENGTH would bring an overall improvement over the baseline, but since – naturally – there are acceptances and rejections with the same length, a small drop in precision is not unsurprising.

Unfortunately, the results of the TURNS featureset imply that just more utterances were classified as rejections overall. Based on the previous sections, we would have hoped that the context-related effects we observed have more discerning strength.

Nevertheless, used together, the featuresets result in a high recall and mitigate the stark decline in precision the featureset TURNS caused. Given that our goal was a high recall this is a desirable effect, but it is unfortunate that we stay below baseline in precision.

[2] Where **n't* is to be read as a wildcard expanding to, *e.g.*, *don't*, *can't*, *wouldn't*.

6 Conclusion and Future Work

We have found noticeable differences in the contextual structure of dialogue surrounding acceptances and rejections. The data we gathered on the Switchboard corpus at least qualitatively supports our theoretical hypotheses. Experimental verification of these differences was hindered by the sparse data we have available, but the contextual features we extracted allow us to identify rejections with a higher precision.

The main method to improve on our results would to use a richer dataset. Unfortunately, our main features, *viz.*, disfluencies, turn length and backchannels, are unique to spoken dialogue and it is hard to come by rich corpora of transcribed and tagged dialogue. Overall, the conversations in the Switchboard corpus are too casual and polite; a corpus that contains more discussion-driven conversation might allow us to extract more useful data. The ICSI Meeting Corpus contains more debate (and thusly more rejections), so it may be more suited to this kind of analysis, *cf.*, (Galley et al., 2004).

Section 4.3 indicates that data sparseness is less of a problem than data spread, since the high standard deviations also occur in the (statistically large enough) set of acceptances. The data might be improved by isolating different types of acceptance and rejection and comparing these with each other, *cf.*, (Misra and Walker, 2013) and (Walker, 1996).

Our experimental results could be improved by using a more sophisticated classifier to train. Most of the features we extracted are read as "per utterance" or "per word" and are therefore not discrete and thereby ill-suited for a Bayesian classifier. An appropriate Vector Space Model might be more useful as a classifying tool. Furthermore, we believe that a classifier that allows for weighting might yield an improvement; it seems reasonable that the contextual features we isolated are less important than the utterance-internal features.

In conclusion, our main hypothesis – that structural and contextual cues can help discerning agreement from rejection – is supported by the data, though the data leaves much to be desired. More sophisticated methods and better training data is required to apply our observations.

References

Or Biran and Owen Rambow. 2011. Identifying justifications in written dialogs by classifying text as argumentative. *International Journal of Semantic Computing*, 5(04):363–381.

D. Bousfield. 2008. *Impoliteness in Interaction*. Pragmatics & beyond. John Benjamins Pub.

Konstantinos Bousmalis, Marc Mehu, and Maja Pantic. 2009. Spotting agreement and disagreement: A survey of nonverbal audiovisual cues and tools. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–9. IEEE.

Donna K Byron and Peter A Heeman. 1997. Discourse marker use in task-oriented spoken dialog. In *Proceedings of Eurospeech*, volume 97, pages 2223–2226. Citeseer.

Raquel Fernandez. 2013. Dialogue. In R. Mitkov, editor, *Oxford Handbook of Computational Linguistics*. Oxford University Press. 2nd Edition.

Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 669. Association for Computational Linguistics.

John J. Godfrey and Edward Holliman. 1997. Switchboard-1 release 2. Linguistic Data Consortium, Philadelphia.

Dustin Hillard, Mari Ostendorf, and Elizabeth Shriberg. 2003. Detection of agreement vs. disagreement in meetings: training with unlabeled data. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003–short papers - Volume 2*, NAACL-Short '03, pages 34–36, Stroudsburg, PA, USA. Association for Computational Linguistics.

Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. 2003. The icsi meeting corpus. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–364. IEEE.

M. McTear. 2004. *Spoken Dialogue Technology: Towards the Conversational User Interface*. Springer.

Amita Misra and Marylin A Walker. 2013. Topic independent identification of agreement and disagreement in social media dialogue. *Evolution*, 460(460):920.

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

scikit-learn developers. 2013. Scikit-learn Documentation version 0.14: 1.7. Naive Bayes. `http://scikit-learn.org/stable/modules/naive_bayes.html`. Accessed: 2013-10-27.

Robert C. Stalnaker. 2013. 7. Assertion. *The Semantics-Pragmatics Boundary in Philosophy*, page 179.

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.

Marilyn A Walker. 1996. Inferring acceptance and rejection in dialog by default rules of inference. *Language and Speech*, 39(2-3):265–304.