Identifying rhetorical questions is relevant for many NLP tasks, including information extraction and text summarization. We present a simple n-gram based language model to classify rhetorical questions in the Switchboard Corpus. We find that incorporating contextual cues achieves the highest performance.

**Features for Identifying Rhetorical Questions**

- Unigrams and bigrams to capture speaker intent
- Linguistic cues namely, strong negative polarity items (NPIs), modal auxiliaries, and certain expressions such as yet and after all
- POS n-grams to capture predictive common grammatical relations
- Contextual cues i.e., phrases uttered prior to the questions and the response to the questions which reflect discourse patterns.

Example conversation:

"... you give them an F on something that does not seem bad to me. What are you telling that student? You’re telling them that hey, you might as well forget it, you know."

If we only consider the question in the above example, it could be a regular question. However, if we consider the precedent and subsequent utterances, we know that it is a rhetorical question.

Accordingly, we consider the following feature sets:

- Unigrams
- Bigrams
- POS bigrams
- POS trigrams

**Result & Analysis**

Table 1 shows the performance of the feature sets cross-validated and trained on 5960 questions in the corpus and tested on 2555 remaining questions. We use $F_1$-score as our main metric due to the skewed nature of our dataset (~5% of test questions rhetorical).

Our results largely reflect our intuition on the expected utility of our various feature sets. While features in the question group prove the most useful single source, the subsequent features outperform precedent features.

Our results suggest that incorporating contextual cues from both directly before and after the question itself outperforms classifiers trained on a naive question-only feature space.

**Conclusion**

Overall, adding in n-gram features from the context improves the performance. We achieve a 53.71% $F_1$-score by adding features extracted from the preceding and the subsequent utterances, which is about a 10% improvement from a baseline classifier using only the features from the question itself.