



Where are we in discourse relation recognition?

Katherine Atwell, Junyi Jessy Li, Malihe Alikhani

Importance of discourse

Discourse analysis is crucial towards comprehending the meaning of texts

For instance, take this example:

“While this book is totally different from any other book he has written to date, it did not disappoint me at all.”

Overall sentiment: **positive**

Limitations of discourse models

Several tasks have found discourse features to be useful:

- Machine comprehension
- Summarization
- Sentiment analysis

However, few works have found discourse *relations* as key features. Possible reasons:

- Discourse parsing is difficult, and parsers are still low-performing
- Parsers are often used on texts outside of the training domain

Our goal

Study the performance of parsers on several different linguistic distributions

- Inter-sentential vs. intra-sentential
- Wall Street Journal vs. other news texts vs. domains outside of news

Identify where parsers are having issues and why

Discourse Frameworks

Penn Discourse Treebank (PDTB)

“While this book is totally different from any other book he has written to date, it did not disappoint me at all.”

Explicit – Contingency.Cause

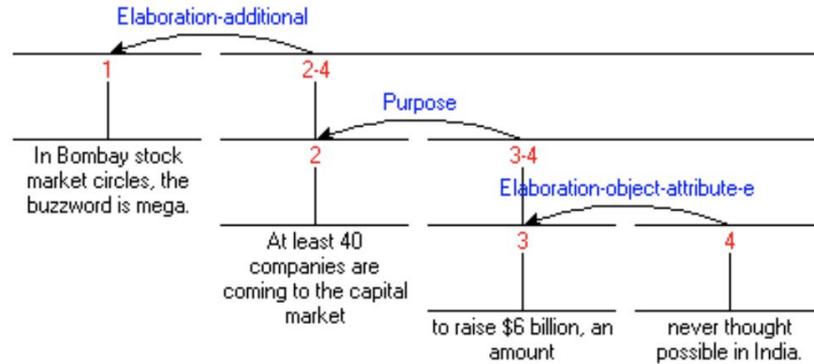
A Lorillard spokeswoman said, *“This is an old story. [In fact] We’re talking about years ago before anyone heard of asbestos having any questionable properties.”*

Implicit – Expansion.Restatement

PDTB-3: contains intra-sentential implicit discourse relations

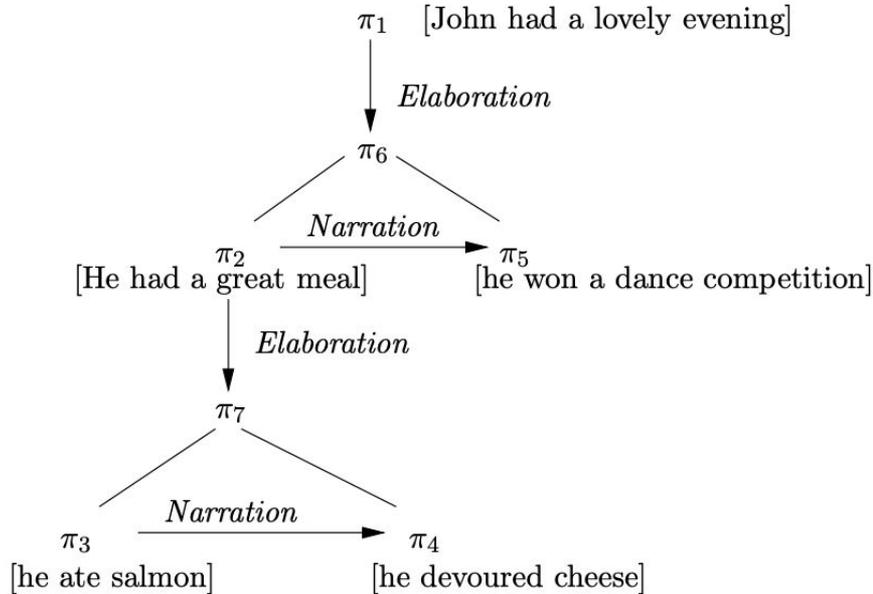
Discourse Frameworks

RST Discourse Treebank



Discourse Frameworks

Segmented Discourse Representation Theory (SDRT)



Datasets

WSJ articles

Wall Street Journal articles contained in the COHA corpus

Other news articles

Non-WSJ articles contained in the COHA corpus

GUM corpus

Contains gold RST labels for texts of 8 different genres: Academic, Biography, Fiction, Interview, News, How-To, Travel, Reddit

Distributional Shifts

Inter-sentential vs. intra-sentential

- Inter- vs. intra-sentential relations in RST and PDTB

Text domain

- WSJ articles, other news articles, GUM corpus

Discourse Parsing Models

Penn Discourse Treebank

- Wang and Lan (2015) end-to-end parser
- DiscoEval BERT-based model

RST Discourse Treebank

- Wang (2017) parser

Manual Annotation

Annotators

- Trained on the PDTB-2 annotation manual
- Faculty and student in linguistics departments
- Given only the two arguments in front of them; could refer to the full article text if needed

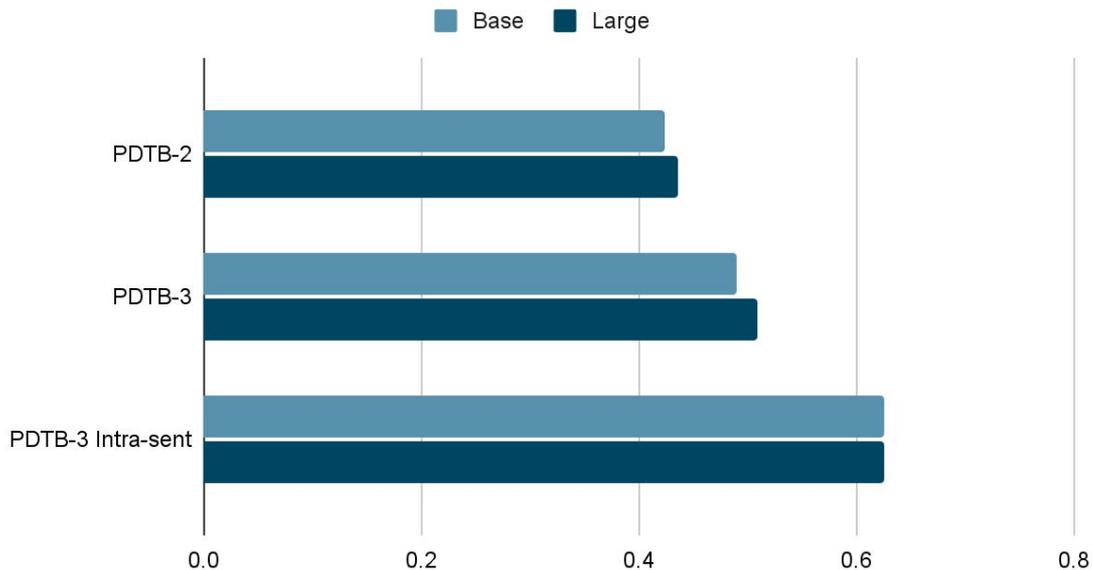
Sample

- Relations were sampled from WSJ, other news articles, and GUM corpus
- Randomly chosen from the relations output by the parser
- 64 implicit relations total (only implicit results were published)

Takeaway I

Transformer-based models perform better on linguistically different intra-sentential relations than they do on inter-sentential relations.

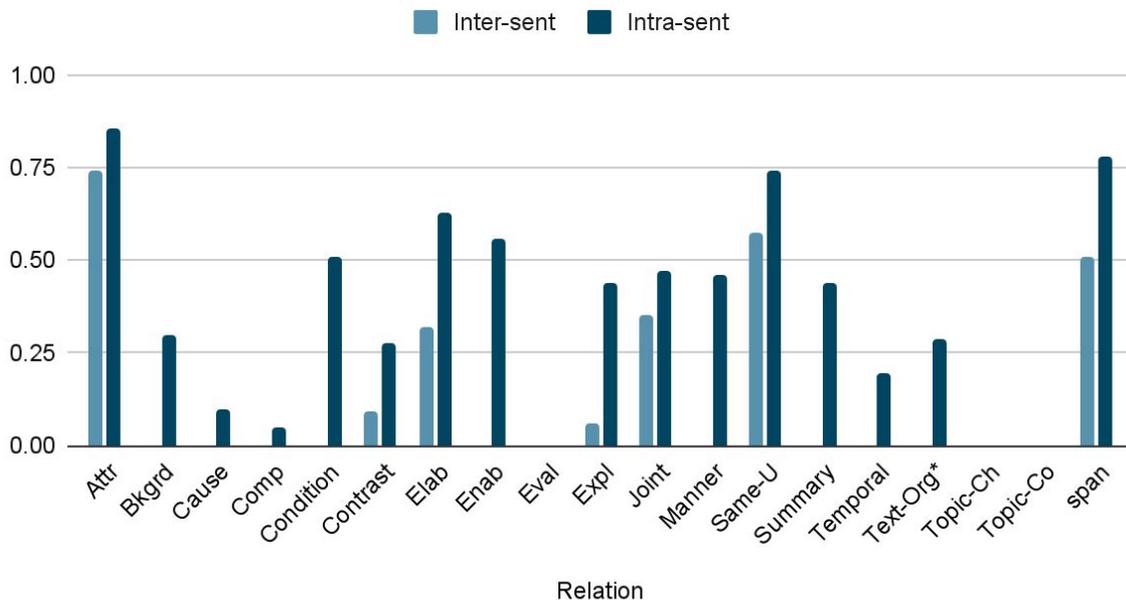
Model Accuracies



Takeaway I

Transformer-based models perform better on linguistically different intra-sentential relations than they do on inter-sentential relations.

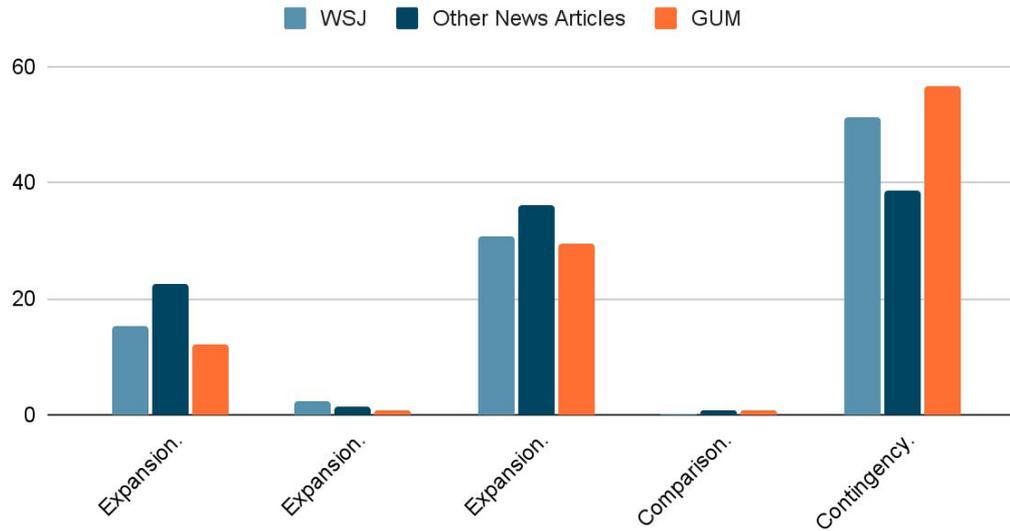
F-1 Scores on each Discourse Relation



Takeaway 2

Parsers struggle to identify implicit relations from less frequent classes

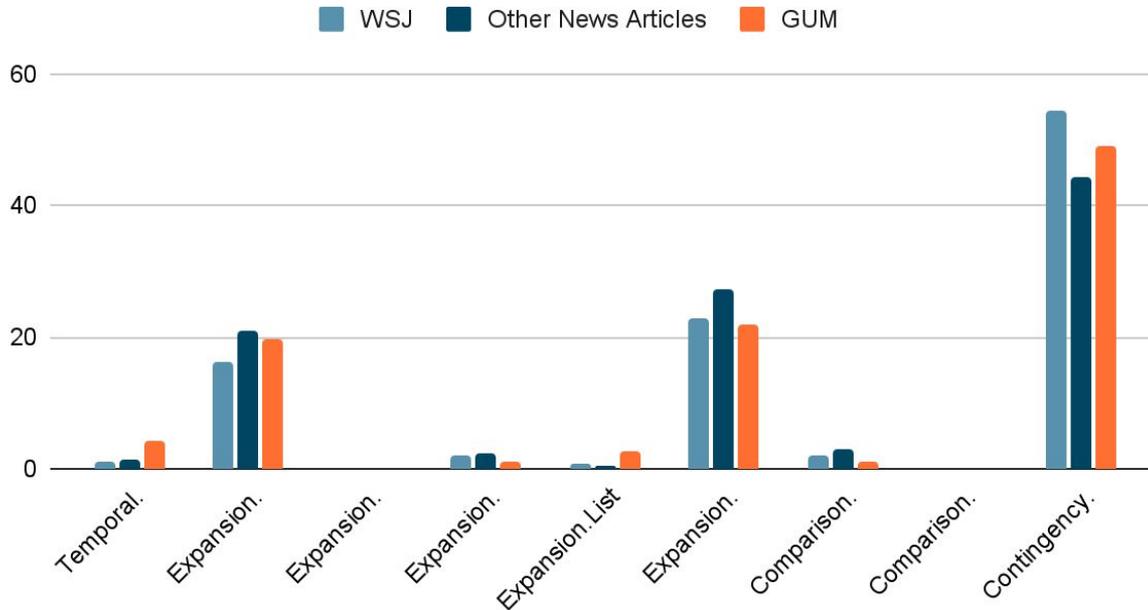
Percentage of times predicted



Takeaway 2

Parsers struggle to identify implicit relations from less frequent classes

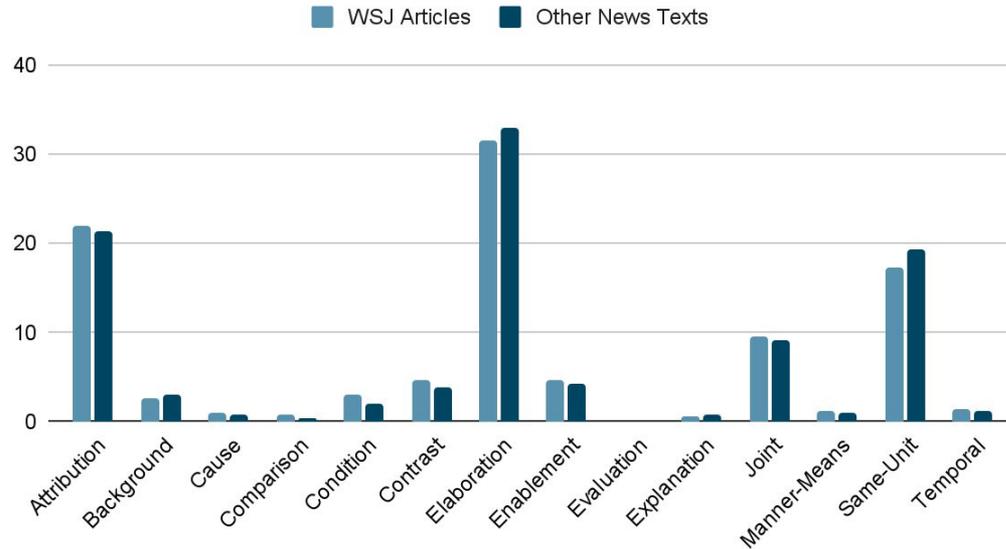
Percentage of times predicted



Takeaway 2

Parsers struggle to identify implicit relations from less frequent classes

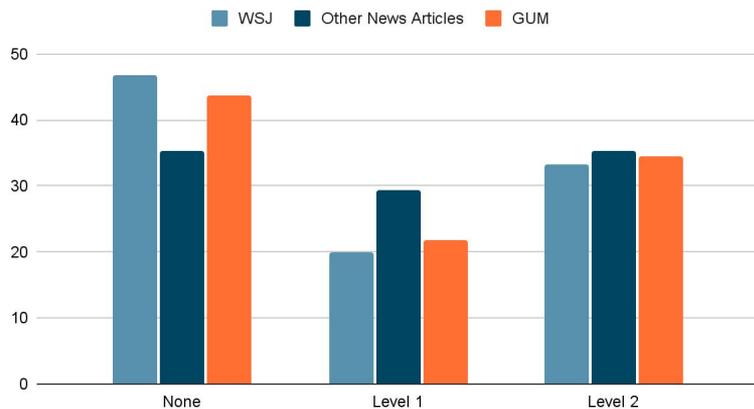
Percentage of times predicted



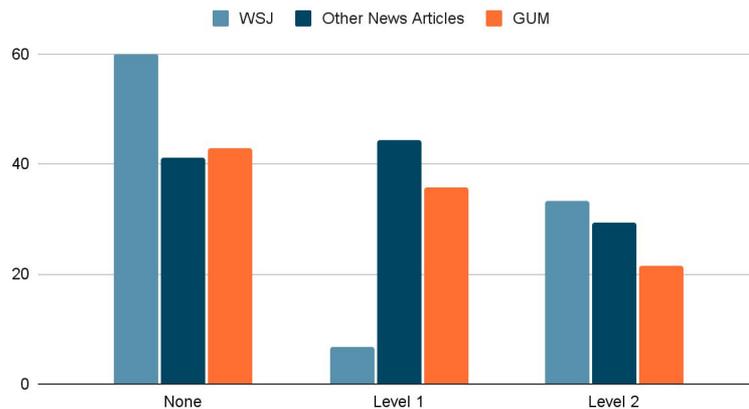
Takeaway 3

Models fail to generalize to both in-domain and out-of-domain data, and different errors are seen for different domains.

Wang et al Accuracies



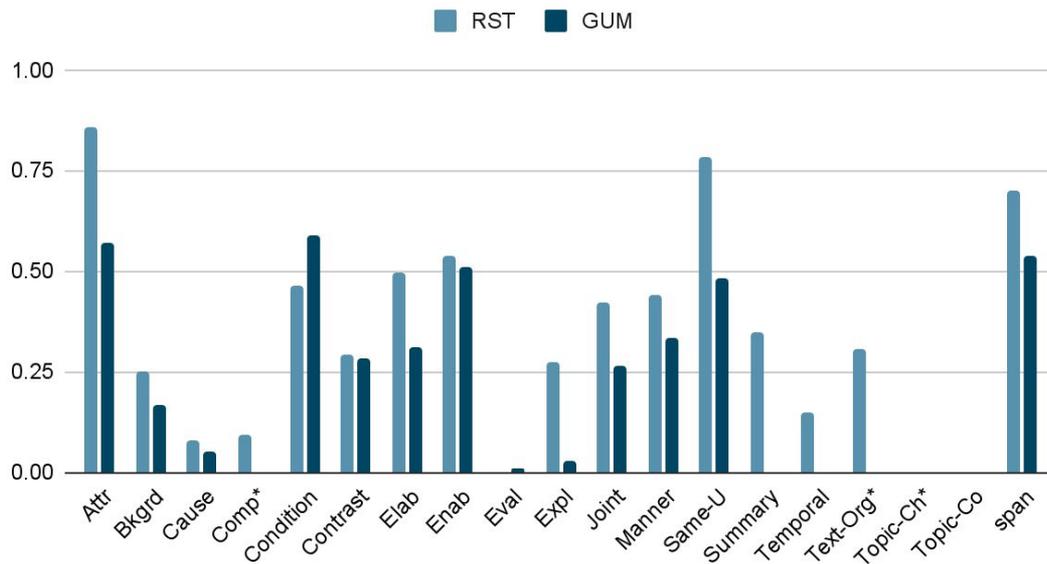
DiscoEval Accuracies



Takeaway 3

Models fail to generalize to both in-domain and out-of-domain data, and different errors are seen for different domains.

F-1 scores for each relation



Insight for future model development

More context than the two arguments is often needed to determine the correct discourse relation

One housewife says : “With an electric kitchen I have to do my whole day’s cooking the day before – and that during a couple of hours, not knowing from one minute to the next what time the power is coming on. ”

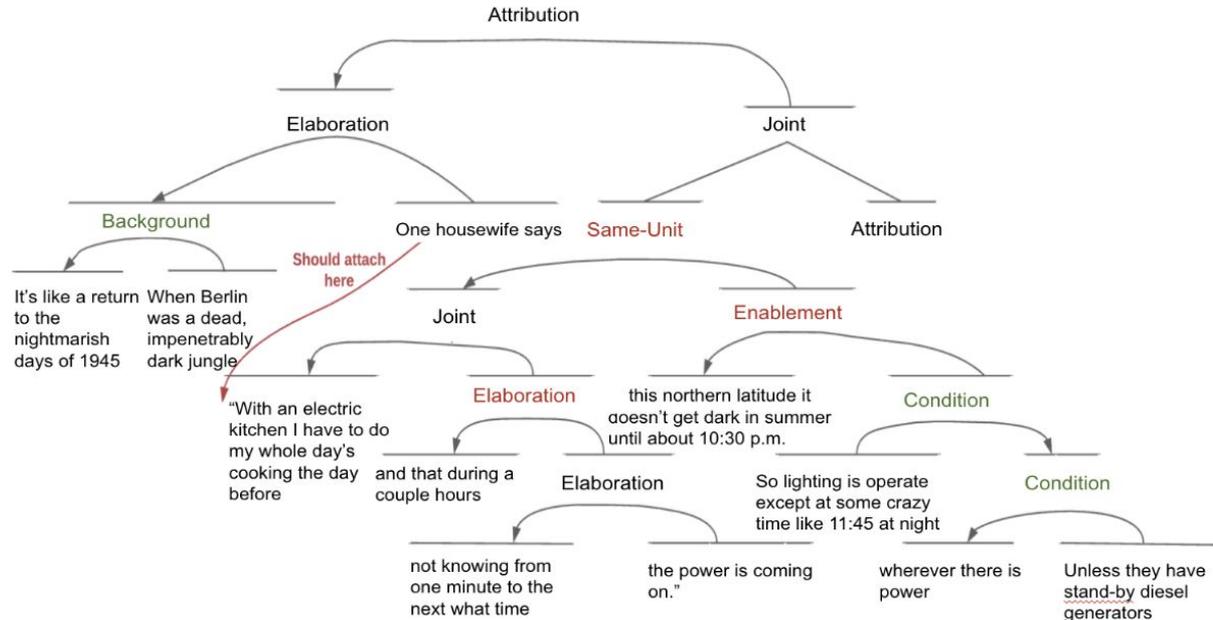
In this northern latitude it doesn’t get dark in summer until about 10:30 p.m. so lighting is operate except at some crazy time like 11:45 at night , whenever there is power, unless they have stand-by diesel generators. There’s a year’s supply of diesel oil here.

Original label: Contingency.Condition

Revised label: Comparison.Contrast

Insight for future model development

Attachment issues tend to occur throughout the RST parse tree, and relations are often misclassified as Same-Unit and Elaboration

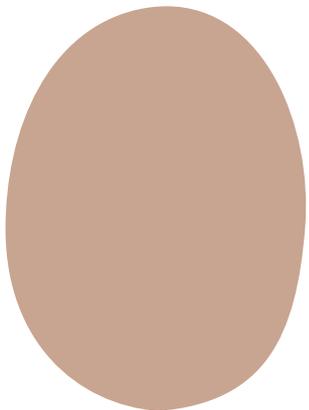
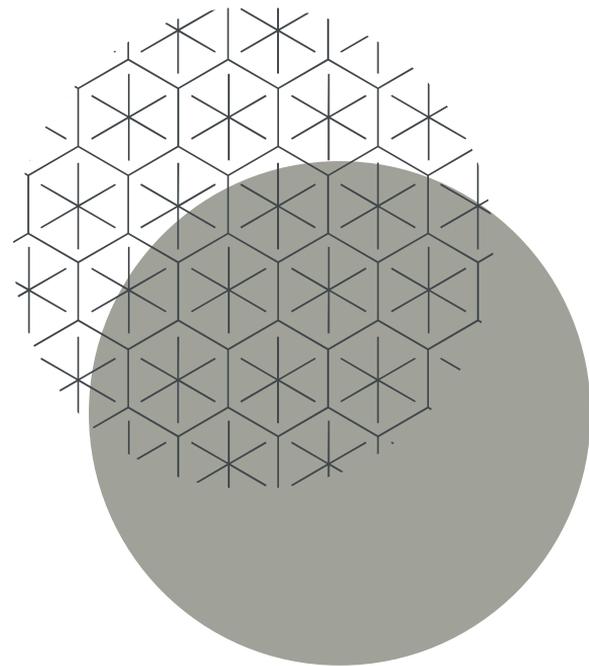
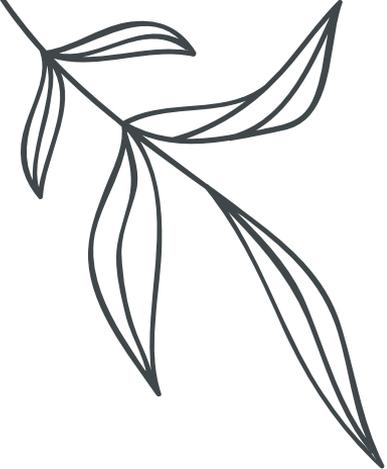


Conclusion

- Models do better at predicting intra-sentential discourse relations compared to inter-sentential discourse relations
- Class imbalance leads to lack of predictions for underrepresented classes
- Discourse parsing models appear to struggle under any kind of domain shift

Important Next Steps

- Address class imbalance issues
- Incorporate more context beyond the two arguments into parsing models
- Unsupervised approaches to classifying discourse relations



Thank you!

Email: kaa139@pitt.edu

Website: <https://katherine-atwell.github.io>

Github: <https://github.com/katherine-atwell>