A Details on data processing

For each dataset, we preprocess each paragraph as follows:

- The sentences which length is less than 5 or higher than 25 are filtered out to remove too short or too long sentences.
- Due to too much noise from News and papers corpus, we make bit aggressive filters. The paragraphs whose last sentence ends with all capital words are removed to filter out the articles with reporter's name or other meta information (e.g., location, press name). Also, paragraphs whose last sentences don't end with sentence-ending marks (e.g., ., !, ?) are also filtered out.
- If any adjacent sentences in a paragraph is identical, we exclude the paragraph. All the duplicate paragraphs are also removed.
- We ignore the paragraph that fails to be parsed by our discourse (i.e. RST) parser. The detail of the parsing would be described in the next section. During the RST parsing, some Stanford dependency parses contain UNK token that is mismatched with our tokenizer (i.e. nltk's word tokenizer). Then, we also ignore such cases (only 1.5% of entire dataset).

B Theoretical difference between linguistic and latent relations

We briefly summarize the fundamental differences of the two relation forms:

- While labels in linguistic relations are interpretable, accuracy of the labels highly depends on the performance of the discourse parser. On the other hand, latent representations with delta operation do not suffer from out-of-domain or accuracy problems that external parsers may bring in.
- Linguistic relations can hold over long distances with many things in between (e.g., Solutionhood), while the latent ones are always immediately adjacent.
- Linguistic ones are fairly coarse-grained and non-continuous, often making them inapplicable to other continuous models (e.g., a neural network), while latent ones are by definition continuous, always making them applicable.
- Linguistic relations are often ambiguous or unclear, while latent ones can easily hybridize and represent two more relations at the same time, at the cost of being indefinable.

Table 5: Example texts produced by different models. Given **FIRST** and **LAST** sentences, the models generate middle sentences (e.g., [M1] → [M2]→ [M3]..). **REF** and **HUMAN** are reference middle sentences and sentences written by human annotator, respectively.

FIRST: Okay, 'Molinari sighed. **LAST**: His voice rose to a shout of anger.
REF: [M1] I'll get up; just leave me alone, will you, for chrissake?' [M2] He stirred about, struggling to get from the bed. [M3] Okay - I'll get up; will that satisfy you?' **HUMAN**: [M1] What? Is this everything you want me to talk to? he couldn't stop laughing this situation. [M2] I don't want to talk with you anymore, Molinari said. [M3] You make me crazy. You disappoint me so badly!!
FT2SEQ: [M1] you are a few times and you are not a good person [M2] she has a good idea and the others are not a bit of pleasure [M3] he was a little bit of his own life **H-FT2SEQ**: [M1] the two of the men are not a UNK [M2] the two of them are not to be able to make a small map of the ship [M3] he had to be a man who had been a man who had been a man who had been a man **FLOWNET (DELTA)**: [M1] he's a very large man, and a man with a very long way [M2] i'm not sure that i'm not going to be able to get rid of it [M3] he had to be a child and a man **FLOWNET (DISC.)**: [M1] you're going to be a good time [M2] it's a lot of the people, and the other worlds is a simple place to be in the universe that [M3] you're not to be a friend