

# Transformer-based Multi-Party Conversation Generation using Dialogue Discourse Acts Planning

Alexander Chernyavskiy and Dmitry Ilvovsky

National Research University Higher School of Economics, Moscow, Russia

## Abstract

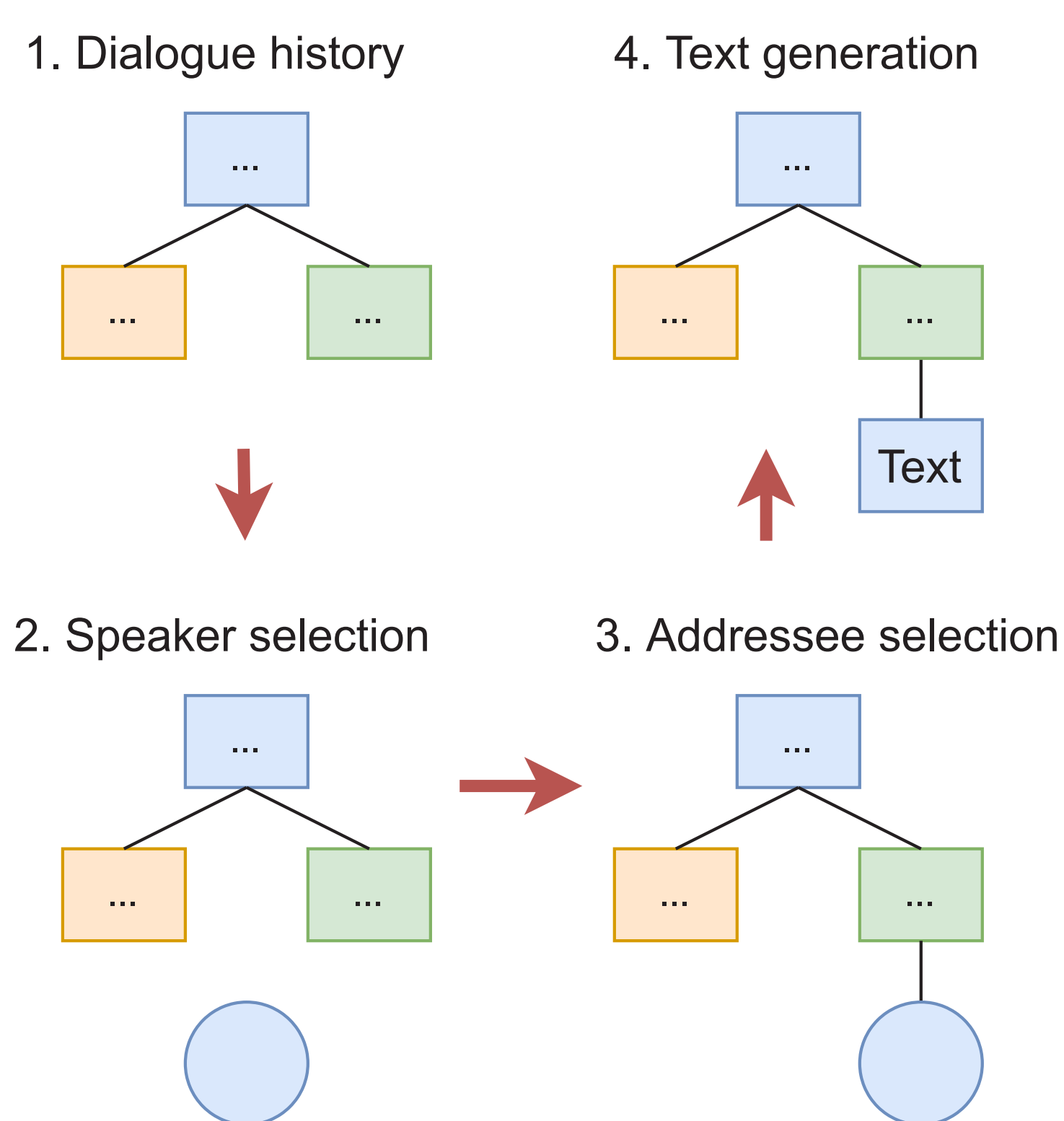
We consider the issue of inconsistent dialogue generation in the case of multi-party conversations (MPC).

- We suggest multi-task learning consisting of dialogue acts planning and response generation to improve the transformer-based MPC generation pipeline.
- We analyze the importance of discourse tokens at the input and at the output.
- We show that the approach converges faster if it has more weight in the loss related to the dialogue acts planning task.

Code:

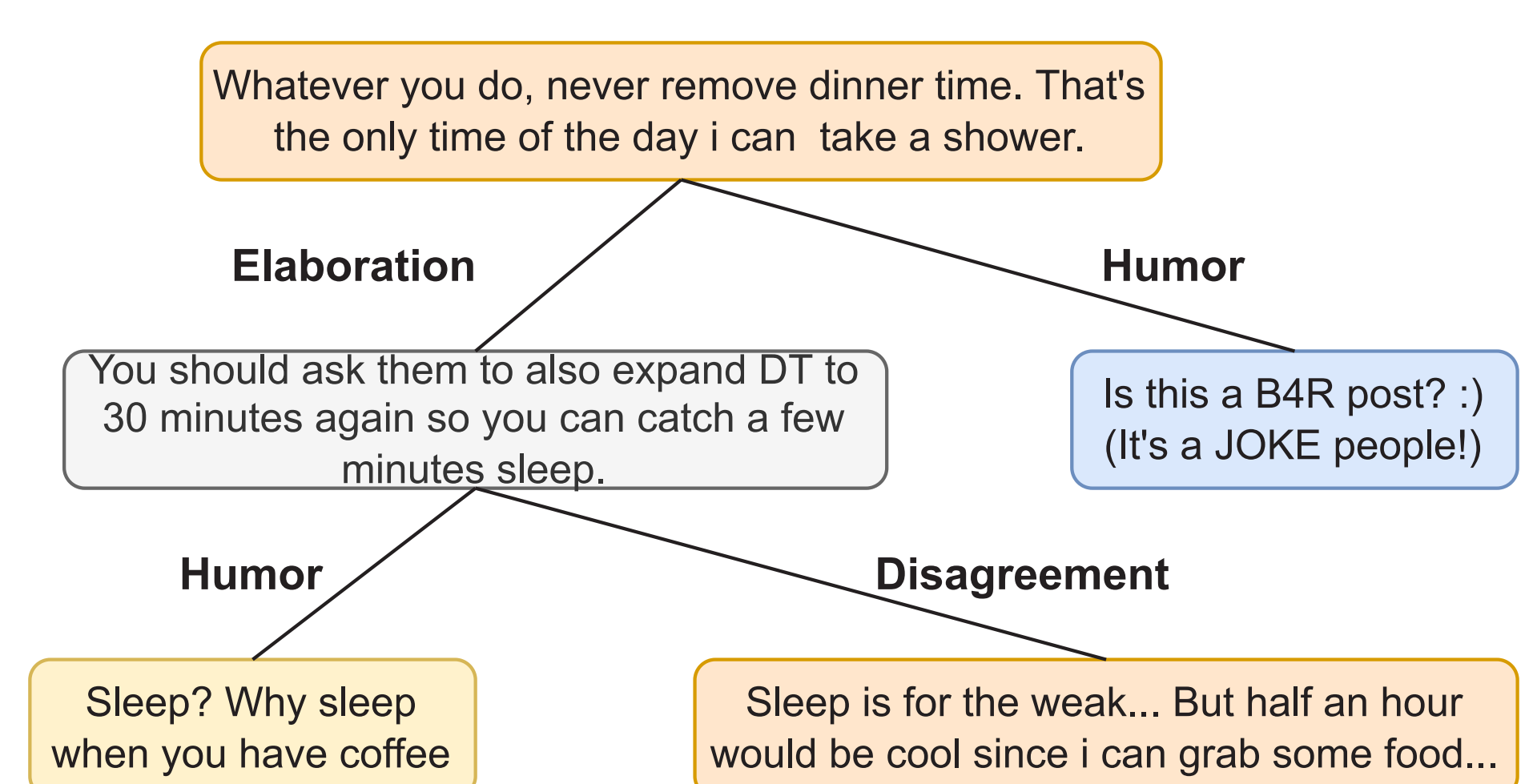
[https://github.com/alchernyavskiy/discourse\\_mpc\\_generation](https://github.com/alchernyavskiy/discourse_mpc_generation)

## Related Work



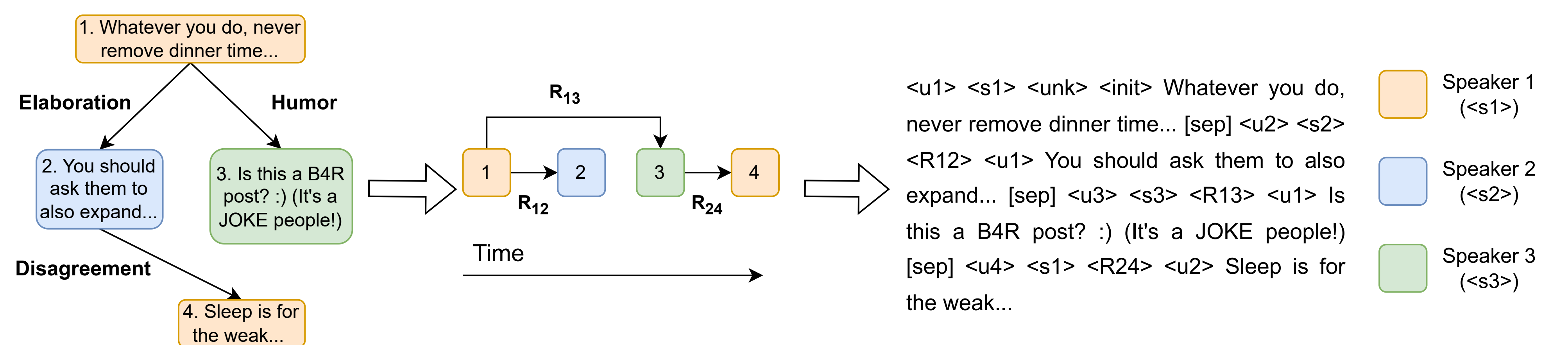
- We consider the **last pipeline's stage** (least explored), associated with the response generation task.
- **Transformers** effectively solve AMR-to-text generation tasks, therefore we used it to **encode and plan discourse structure**.
- Discourse has been explored only in the context of summarization and MRC before.

## Dialogue Acts



- Theory of dialogue acts shows by which **discourse rhetorical relations** (dialogue acts) the individual utterances of the dialogue are connected.
- We use **special tokens** to represent it in the linearized representation.

## Method



- ◇ **Linearization** We suggest to use special tokens as the identifiers of speakers and utterances:  $\{\{s_i\}\}$  and  $\{\{u_i\}\}$  correspondingly. To specify an utterance to respond to at the current turn, we add its representation to the end of the dialogue sequence.
- ◇ **Output format** "<relation> response text"
- ◇ **Loss** Weighted cross-entropy loss where the weight  $\alpha$  is used for the dialogue act tokens.

## Datasets

- **Coarse Discourse Sequence Corpus (CDSC)** is the largest open-source manually annotated dataset of dialogue acts in online discussions. It contains ~9K Reddit threads annotated with 10 main discourse act labels.
- **Movie Reddit** is collected by ourselves. It consists of roughly 90k dialogues from the 25 most popular Reddit subthreads discussing movies, series and TV shows. To obtain discourse acts labels, we trained our own discourse parser from scratch based on the CDSC dataset.

## Results

Dataset	Setting	ROUGE-1	ROUGE-2	ROUGE-L	BLEU-1	BLEU-2
CDSC	full discourse data	<b>9.34</b>	<b>0.66</b>	<b>8.37</b>	<b>8.68</b>	<b>0.32</b>
	no discourse in resp.	7.39	0.48	6.64	6.53	0.30
	no discourse at all	7.44	0.43	6.71	6.58	<b>0.32</b>
Reddit	full discourse data	<b>8.89</b>	<b>0.58</b>	<b>7.96</b>	<b>8.11</b>	0.17
	no discourse in resp.	7.76	0.54	7.07	6.80	<b>0.20</b>
	no discourse at all	7.45	0.51	6.77	6.47	0.17

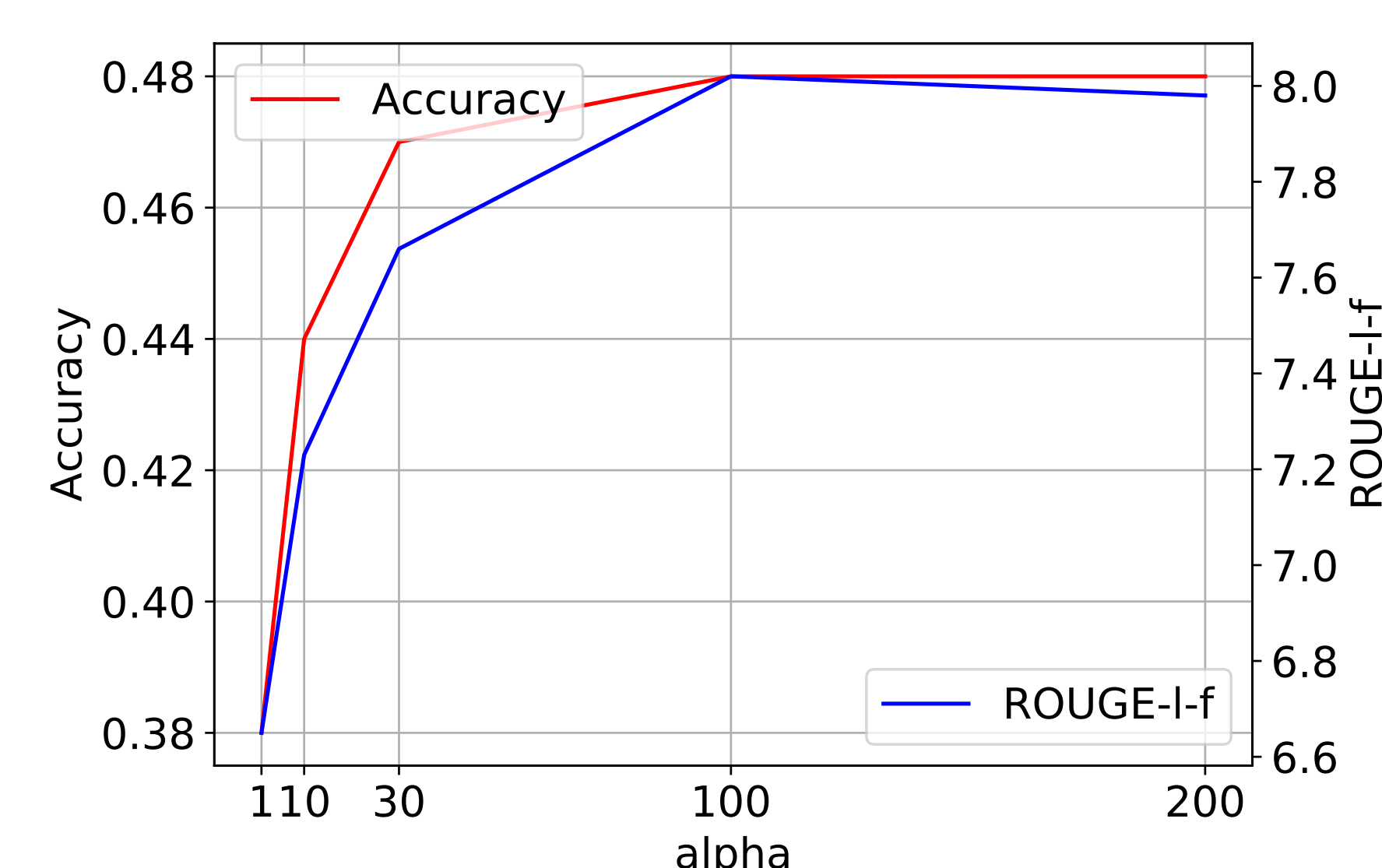
### Automatic evaluation:

- ◇ Automatic evaluation for BART (above) and T5 indicates that discourse planning simplifies response generation.
- ◇ The model featuring a discourse planning performs significantly better even considering the automatically labeled data.

### Human evaluation:

- ◇ 200 random dialogues from the CDSC test dataset; three criteria with 0-2 scales
- ◇ responses produced by the custom approach are preferable in more cases
- ◇ discourse-based model's responses are more coherent despite perhaps less semantically appropriate formulations

Model	# better	Coherence	Meaning.
Base	62	1.11	1.32
Disco	<b>83</b>	<b>1.32</b>	<b>1.33</b>



### Convergence speed:

- ◇ We estimate the fewest number of steps to get high quality (calculated quality after 2 epochs).
- ◇ Strong correlation between Accuracy and ROUGE values, suggesting that improved discourse planning improves the overall quality of language modeling.
- ◇ The approach converges faster if the discourse planning task has more weight.

### Findings:

- ◇ The confusion matrices show that the custom model successfully **plans not only common relations**, such as Elaboration and Answer. It more accurately determines Appreciation, Question and even such relations as Disagreement and Humor.
- ◇ Accurate choice of a dialogue act at the start almost always results in valid responses.