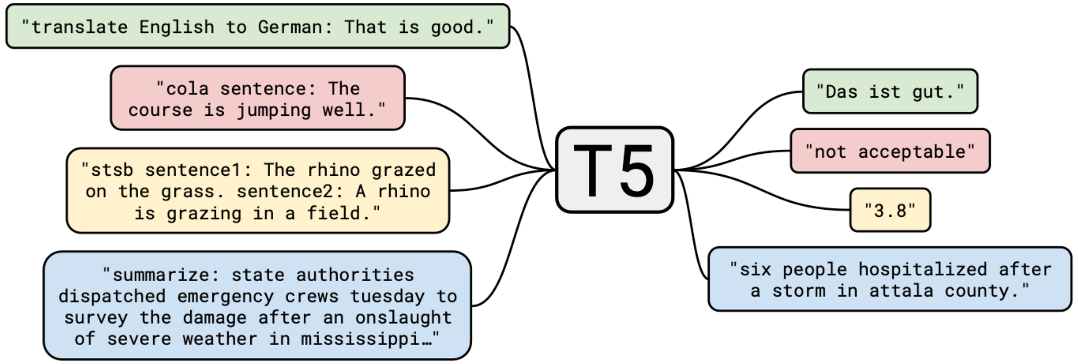
Dialogue Discourse Parsing as Generation: a Seq-to-Seq LLM-based Approach

Chuyuan Li, Yuwei Yin, Giuseppe Carenini

University of British Columbia

SIGdial 2024, September 7, Kyoto University

In the age of Large Language Models



Encoder-decoder Models

- T5
- Flan-T5
- T0
- ...

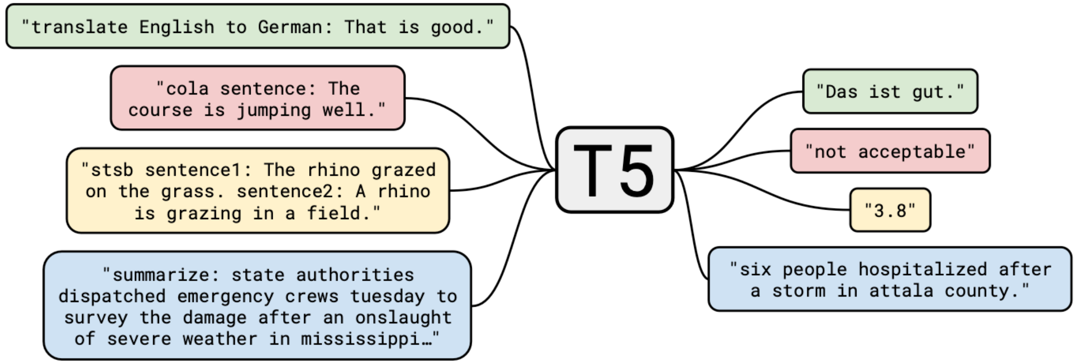
Decoder-only Models

- GPT-3.5, GPT-4
- Llama2, Llama-3.1
- Mistral
- ...

Remarkable ability

Text understanding
Generation
Coding
Reasoning
...

In the age of Large Language Models



MISTRAL AI_

Claude 3

Our Research Goal:

Leverage LLMs for discourse structure prediction without (as far as possible) explicitly designing parsing modules or changing the architecture of LLMs.

Our Approach:

Turning parsing task into a seq2seq generation task, so that we can leverage latent knowledge captured by powerful LLMs.

In the age of Large Language Models

Inspired by the promising results in other structure prediction tasks, e.g., coreference resolution, semantic parsing, etc.

In this paper, we tackle the challenging Discourse Parsing task **with LLMs.**

Don't Parse, Generate! A Sequence to Sequence Architecture for Task-Oriented Semantic Parsing IW3C2 2021

Subendhu Rongali*
University of Massachusetts Amherst
Amherst, MA, USA

Luca Soldaini
Amazon Alexa Search
Manhattan Beach, CA, USA

Seq2seq is All You Need for Coreference Resolution EMNLP 2023

Wenzheng Zhang[1] Sam Wiseman[2] Karl Stratos[1]
[1] Rutgers University [2] Duke University

STRUCTURED PREDICTION AS TRANSLATION BETWEEN AUGMENTED NATURAL LANGUAGES ICLR 2021

Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cicero Nogueira dos Santos, Bing Xiang, Stefano Soatto
Amazon Web Services

Unleashing the True Potential of Sequence-to-Sequence Models for Sequence Tagging and Structure Parsing TACL 2023

Han He
Department of Computer Science
Emory University
Atlanta, GA 30322 USA

Jinho D. Choi
Department of Computer Science
Emory University
Atlanta, GA 30322 USA

How to turn discourse parsing into sequence generation?

Input: sequence of utterances

e1: How do people know about the game?

e2: I did the trials.

e3: I know about it from my boyfriend.

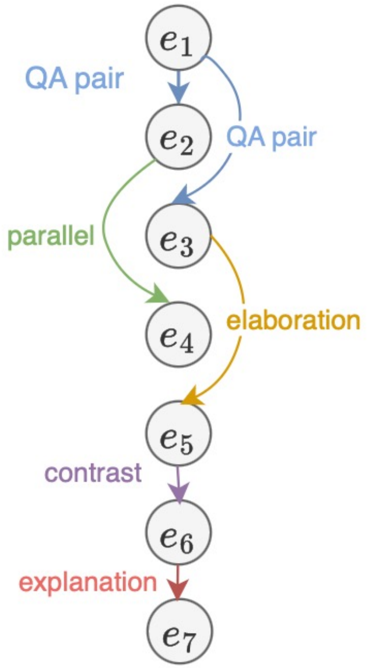
e4: Yeah me too.

e5: I did not do the trials.

e6: I did them,

e7: because a friend did.

Output: graph-like structure



SDRT-style discourse parsing

(Segmented Discourse Representation Theory)

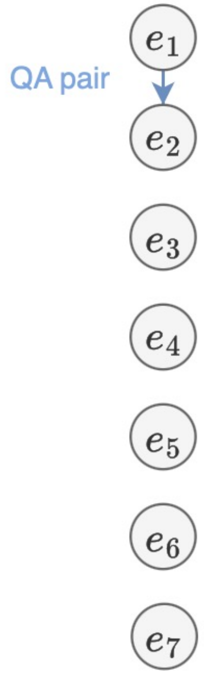
How to turn discourse parsing into sequence generation?

Input: sequence of utterances

e1: How do people know about the game?
e2: I did the trials.
e3: I know about it from my boyfriend.
e4: Yeah me too.
e5: I did not do the trials.
e6: I did them,
e7: because a friend did.

SDRT-style discourse parsing

Output: graph-like structure



Structure linearization

As sequence of triples

(e1, e2, QA pair)

How to turn discourse parsing into sequence generation?

Input: sequence of utterances

e1: How do people know about the game?

e2: I did the trials.

e3: I know about it from my boyfriend.

e4: Yeah me too.

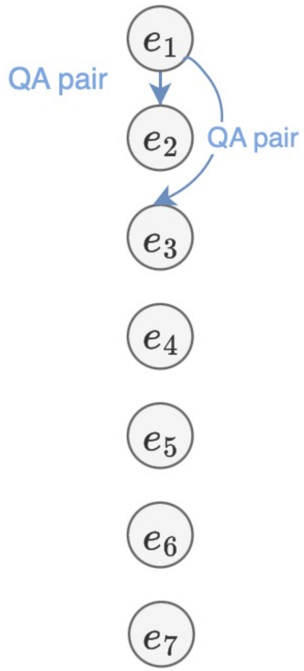
e5: I did not do the trials.

e6: I did them,

e7: because a friend did.

SDRT-style discourse parsing

Output: graph-like structure



Structure linearization

As sequence of triples

(e1, e2, QA pair)

(e1, e3, QA pair)

How to turn discourse parsing into sequence generation?

Input: sequence of utterances

e1: How do people know about the game?

e2: I did the trials.

e3: I know about it from my boyfriend.

e4: Yeah me too.

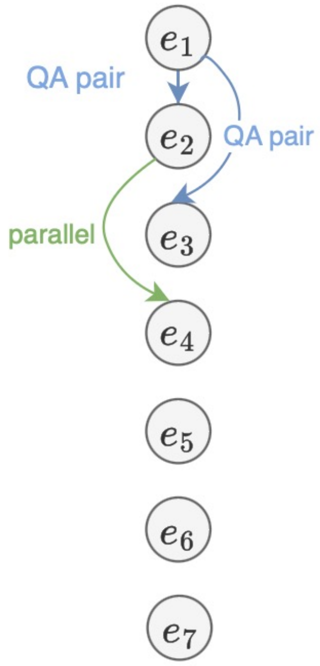
e5: I did not do the trials.

e6: I did them,

e7: because a friend did.

SDRT-style discourse parsing

Output: graph-like structure



Structure linearization

As sequence of triples

(e1, e2, QA pair)

(e1, e3, QA pair)

(e2, e4, Parallel)

How to turn discourse parsing into sequence generation?

Input: sequence of utterances

e1: How do people know about the game?

e2: I did the trials.

e3: I know about it from my boyfriend.

e4: Yeah me too.

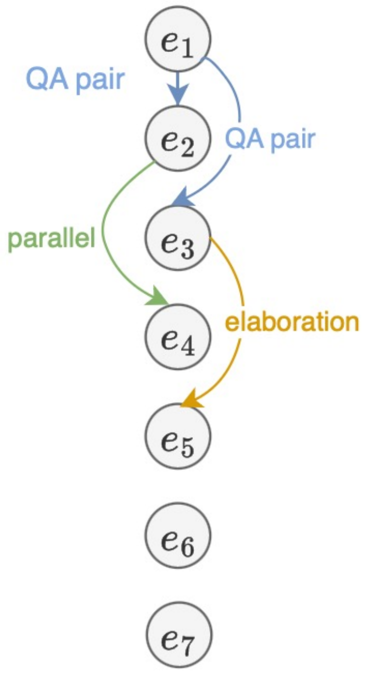
e5: I did not do the trials.

e6: I did them,

e7: because a friend did.

SDRT-style discourse parsing

Output: graph-like structure



Structure linearization

As sequence of triples

(e1, e2, QA pair)

(e1, e3, QA pair)

(e2, e4, Parallel)

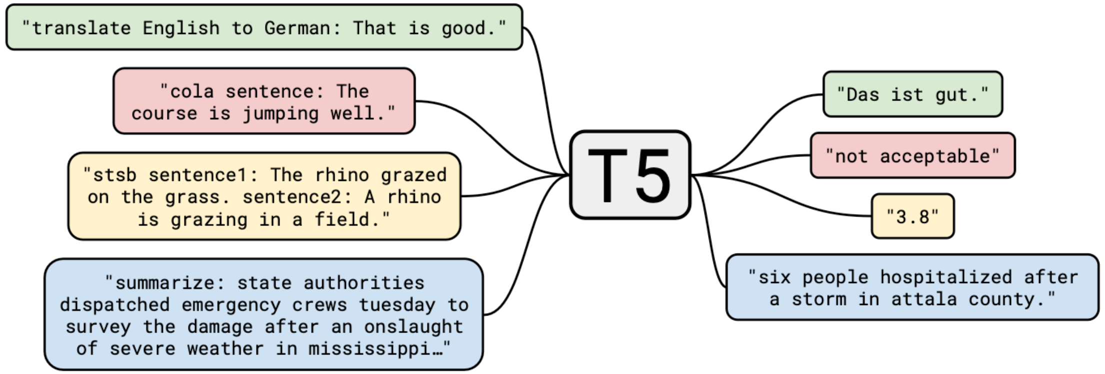
(e3, e5, Elaboration)

...

Outline

- Choice of LLM
- Dialogue Discourse Parsing (DDP) and Seq2Seq Modeling
  - First approach: Seq2Seq-DDP
  - Second approach: Seq2Seq-DDP + Transition
- Evaluation
- Analysis and Future Work

Choice of LLM



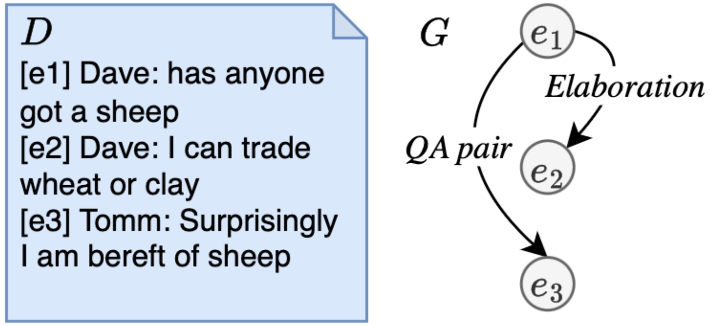
We choose T5 family model: T0

- C4 corpus, 356 billion tokens
- Pretrained on tasks such as multi-doc question answering, natural language inference

- ✓ Good contextual representation for sentence-level reasoning
- ✓ Applied on other structure prediction tasks
  - Coreference resolution [Zhang et al., 2023, Bohnet et al., 2023, Paolini et al., 2021]
  - Semantic parsing [Rongali et al., 2020]
  - Syntactic parsing [He and Choi, 2023]

Discourse Parsing and Seq2Seq Modeling

Discourse parsing



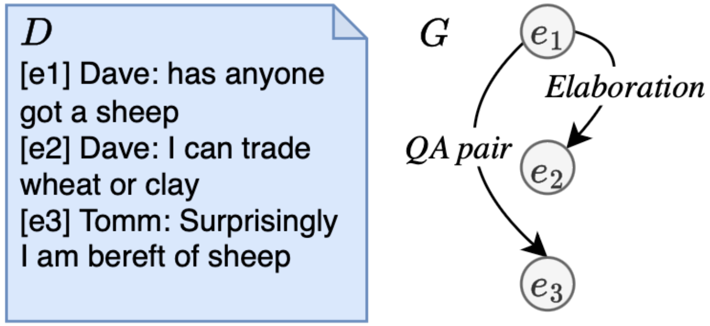
$$D = \{e_1, e_2, e_3\}$$

$$F = \{(e_1, e_2, \text{elaboration}), (e_1, e_3, \text{QA pair})\}$$

$$e : \text{utterances}, \mathbb{R} : \text{relations}$$

Discourse Parsing and Seq2Seq Modeling

Discourse parsing

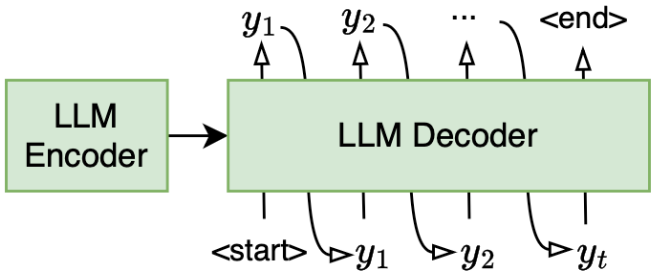


$$D = \{e_1, e_2, e_3\}$$

$$F = \{(e_1, e_2, \text{elaboration}), (e_1, e_3, \text{QA pair})\}$$

$$e : \text{utterances}, \mathbb{R} : \text{relations}$$

Seq2Seq modeling



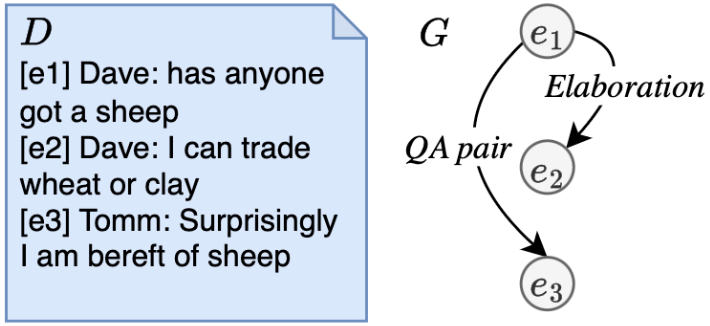
$$x : \text{source sequence}$$

$$y : \text{target sequence}$$

$$p(y|x;\theta) = \prod_{t=1}^{T} p(y_t|y_1, \ldots, y_{t-1}, x; \theta)$$

Discourse Parsing and Seq2Seq Modeling

Discourse parsing



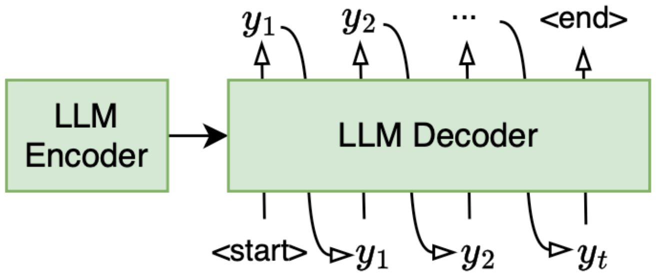
$D = \{e_1, e_2, e_3\}$

$F = \{(e_1, e_2, \text{elaboration}), (e_1, e_3, \text{QA pair})\}$

e : utterances, $\mathbb{R}$: relations

(D, F)

Seq2Seq modeling



x : source sequence

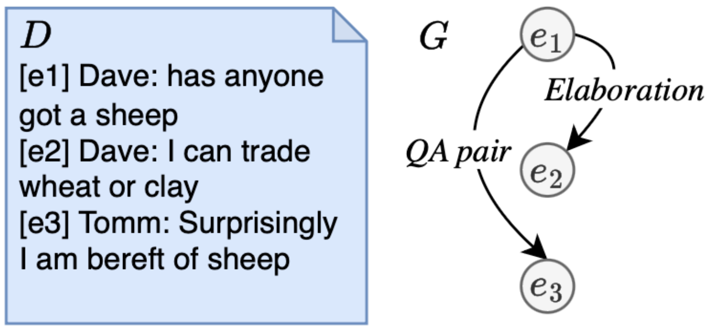
y : target sequence

$$p(y|x;\theta) = \prod_{t=1}^{T} p(y_t|y_1, \ldots, y_{t-1}, x; \theta)$$

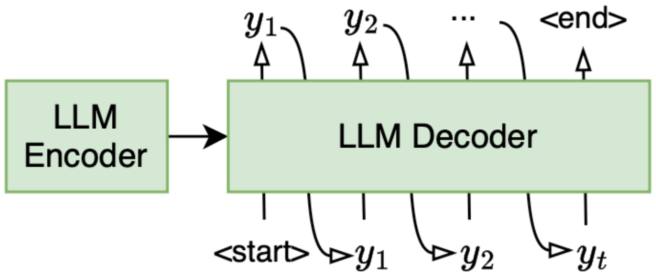
(x, y)

Discourse Parsing and Seq2Seq Modeling

Discourse parsing



Seq2Seq modeling

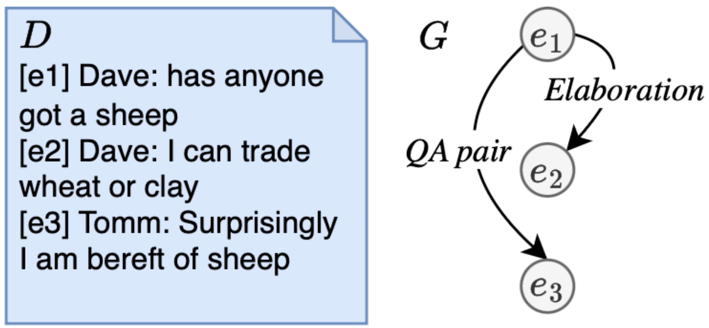


$(\mathcal{D}, F)$ as (x, y)

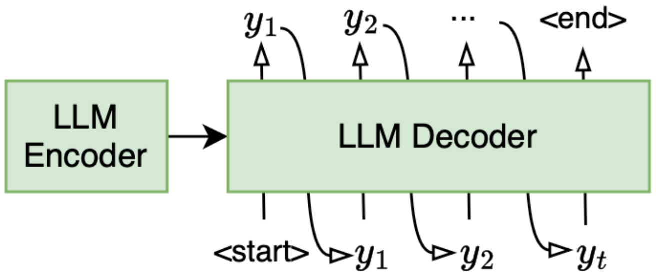
- Translation of **D to x** and **F to y**
  - Straightforward from D to x
  - What about from F to y?
    - → "*Linearization*" process for structured object F

Discourse Parsing and Seq2Seq Modeling

Discourse parsing



Seq2Seq modeling

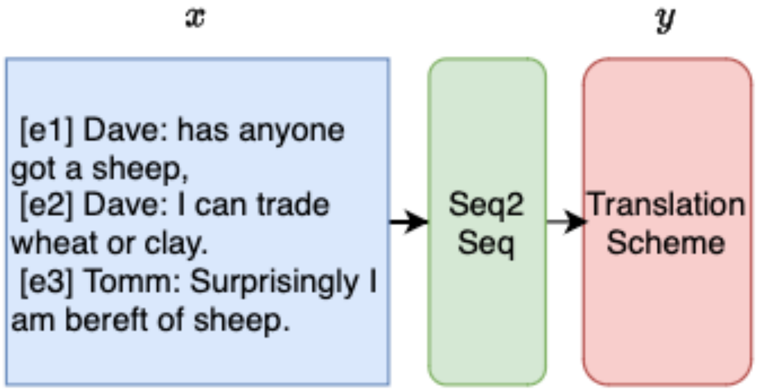


$(\mathcal{D}, F)$ as (x, y)

- Translation of **D to x** and **F to y**
  - Straightforward from D to x
  - What about from F to y?
    - → "*Linearization*" process for structured object F
- Conditional probability **p(y|x)**
  - What is x?
    - The whole document or some utterances?
    - → Two approaches: end-to-end approach and transition-based approach

First approach: Seq2Seq-DDP

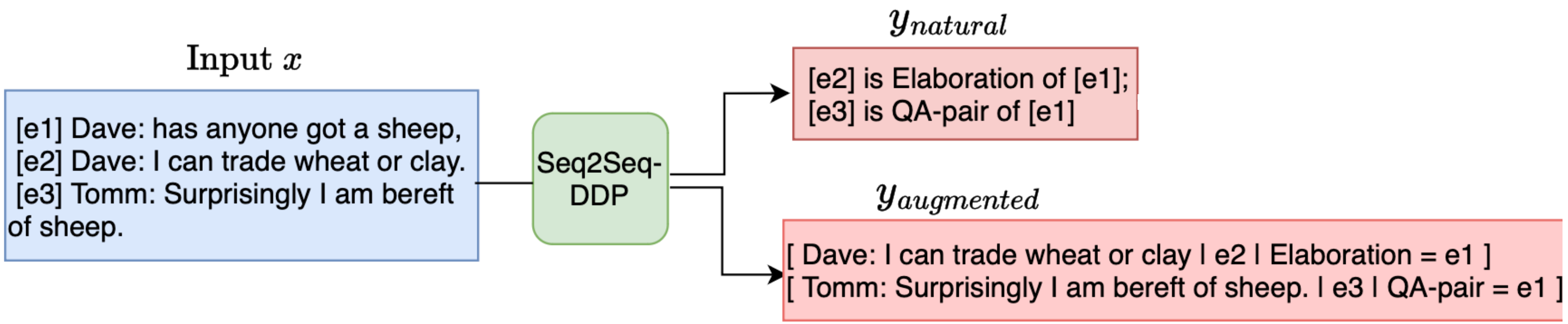
- Y(natural) is a sequence of elements with a structure: $e_i \text{ is } r_{ki} \text{ of } e_k$
  - Close to natural language
  - Use EDU markers to represent utterance
  - *Example for the 1st pair: "e2 is elaboration of e1"*



- Y(augmented) is a sequence of elements with structure: $[\text{ raw text } | e_i | r_{ki} = e_k]$
  - Scheme also used in semantic role labeling and coreference resolution tasks
  - Replicates the input sentence and augments it with EDU marker, link and relation
  - *Example for the 1st pair: "[Dave: I can trade wheat or clay | e2 | Elaboration = e1]"*

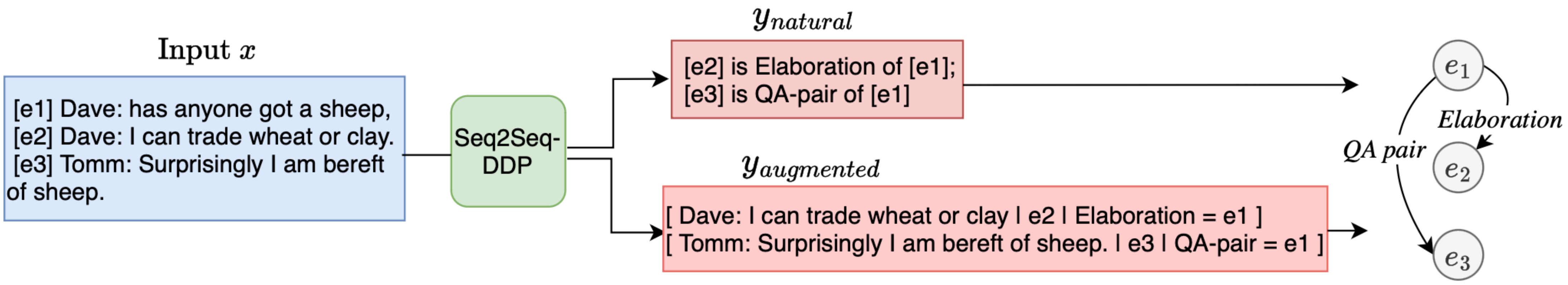
First approach: Seq2Seq-DDP

- Y(natural) is a sequence of elements with a structure: e_i is r_{ki} of e_k
- Y(augmented) is a sequence of elements with structure: $[\,\text{raw text}\,|e_i|r_{ki} = e_k]$
- With a full document x as input, the output looks like:



First approach: Seq2Seq-DDP

- Y(natural) is a sequence of elements with a structure: e_i is r_{ki} of e_k
- Y(augmented) is a sequence of elements with structure: $[\text{ raw text } | e_i | r_{ki} = e_k]$
- With a full document x as input, the output looks like:



- Last step: from Y(natural) and Y(augmented) sequences to the target discourse graph with a simple decoding algorithm.

Analysis of Seq2Sq-DDP Approach

Pros

- Straightforward *linearization* process
- Straightforward conditional probability calculation p(y|x)

  where x = $D = \{e_0, e_1, \ldots e_n\}$

Cons

- Weak supervision in long sequences. The longer the document, the harder it is for the model to retrace previous predictions.
- Consecutive output requires extra attention to some properties such as *counting*, which LLMs struggle with (Kojima et al., 2022).

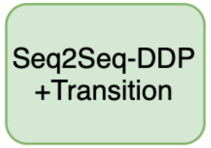
Second approach: Seq2Seq-DDP + Transition

- Related to the *deterministic dependency parsing algorithm* [Nivre, 2003, 2008]
  - Buffer: stores all EDUs
  - States: keeps track of EDU_i being processed
    - Cs: initial state
    - Ct: set of final states
  - Actions: given a state, it defines which *link*(s) and *relation*(s) to assign to EDU_i.

- → Focus on one EDU (utterance) at a time.
- → Prediction is incremental and takes into account the previous states.

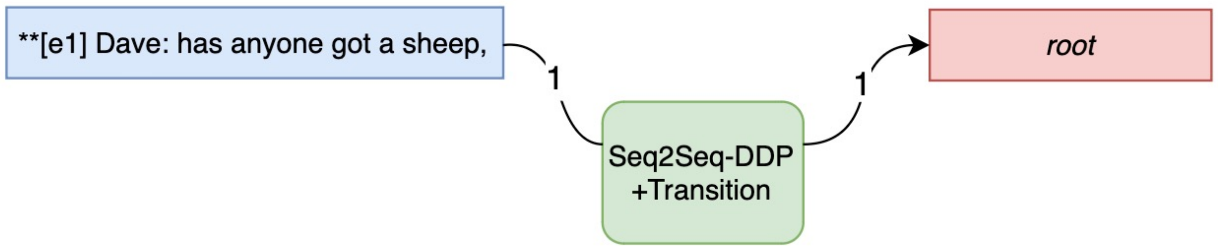
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	...	...
...		
$C_t = \{c \in C \mid c = (e_n, F)\}$		

Second approach: Seq2Seq-DDP + Transition



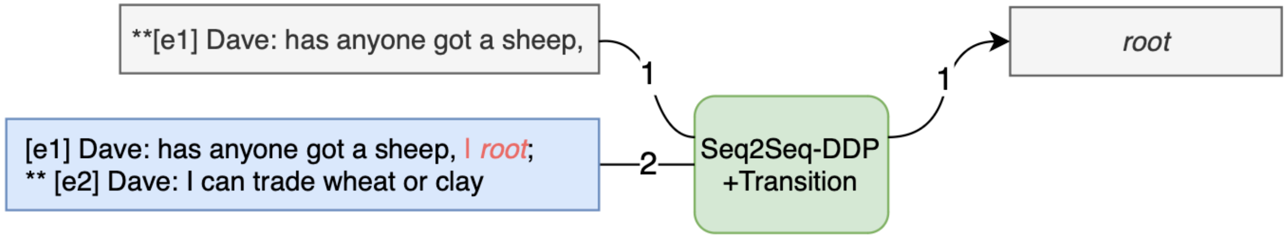
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$		

Second approach: Seq2Seq-DDP + Transition



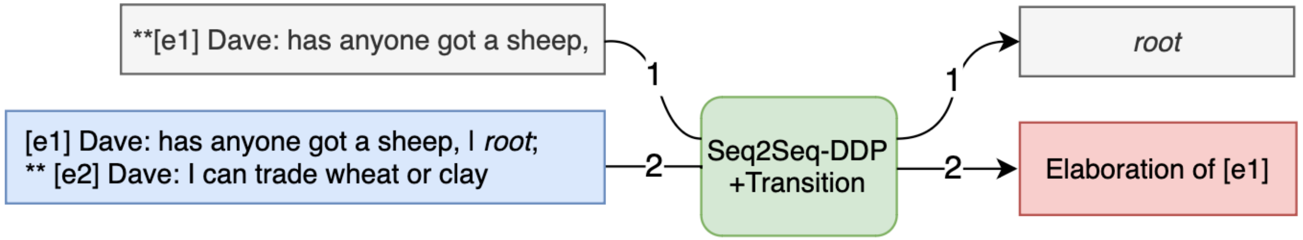
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	(e_1, e_0), root

Second approach: Seq2Seq-DDP + Transition



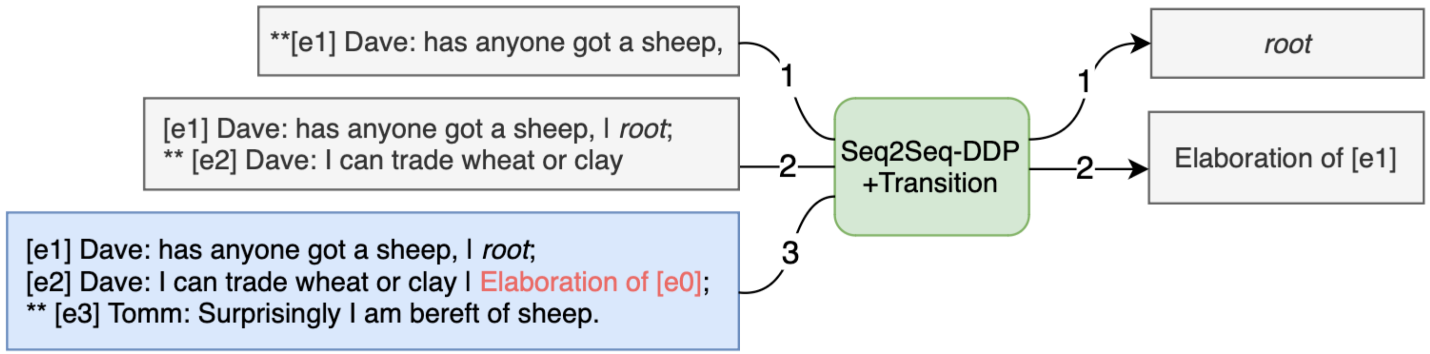
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	(e_1, e_0), root
$c_1 = (e_1, \{(e_1, e_0, \text{root})\})$	e_2	

Second approach: Seq2Seq-DDP + Transition



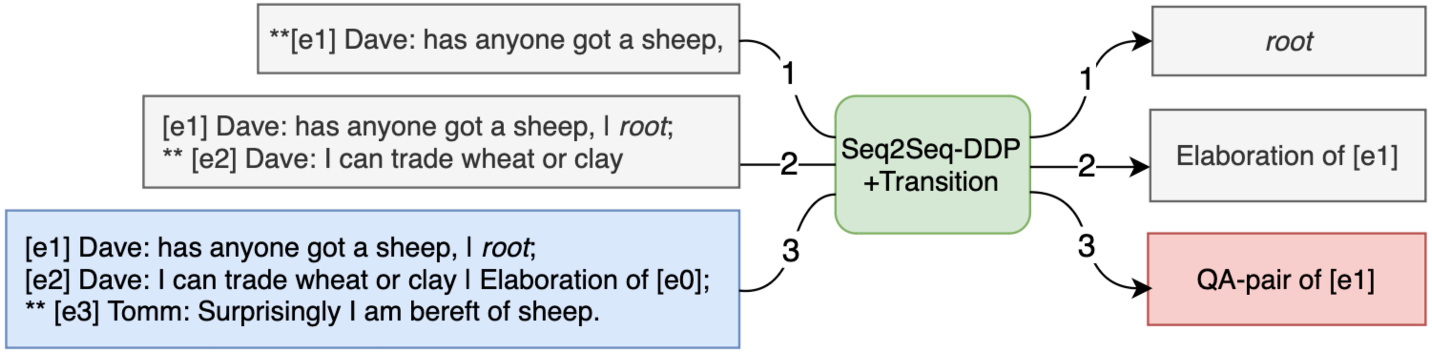
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	$(e_1, e_0), \mathrm{root}$
$c_1 = (e_1, \{(e_1, e_0, \mathrm{root})\})$	e_2	$(e_2, e_1), \mathrm{elaboration}$

Second approach: Seq2Seq-DDP + Transition



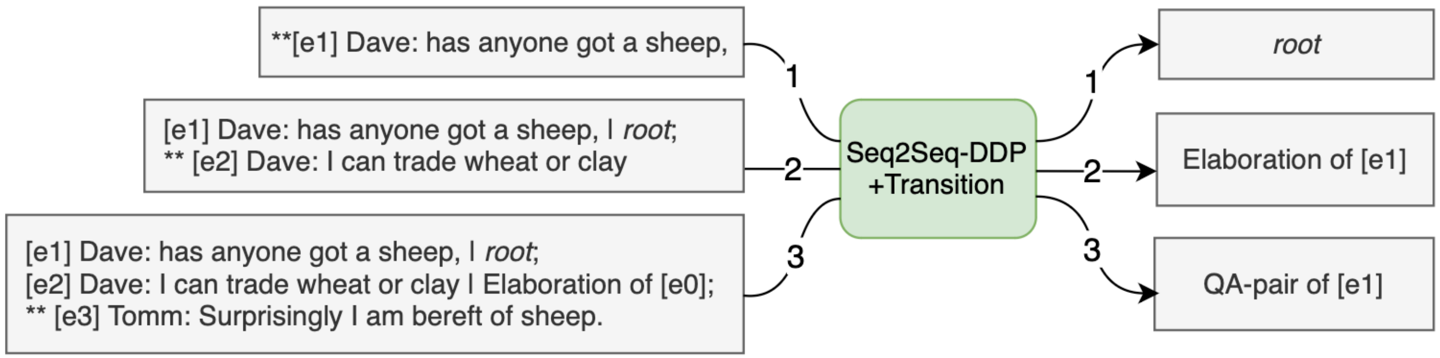
State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	$(e_1, e_0), \mathrm{root}$
$c_1 = (e_1, \{(e_1, e_0, \mathrm{root})\})$	e_2	$(e_2, e_1), \mathrm{elaboration}$
$c_2 = (e_2, \{(e_1, e_0, \mathrm{root}), (e_2, e_1, \mathrm{elaboration})\})$	e_3	

Second approach: Seq2Seq-DDP + Transition



State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	$(e_1, e_0), \mathrm{root}$
$c_1 = (e_1, \{(e_1, e_0, \mathrm{root})\})$	e_2	$(e_2, e_1), \mathrm{elaboration}$
$c_2 = (e_2, \{(e_1, e_0, \mathrm{root}), (e_2, e_1, \mathrm{elaboration})\})$	e_3	$(e_3, e_1), \mathrm{qa\text{-}pair}$

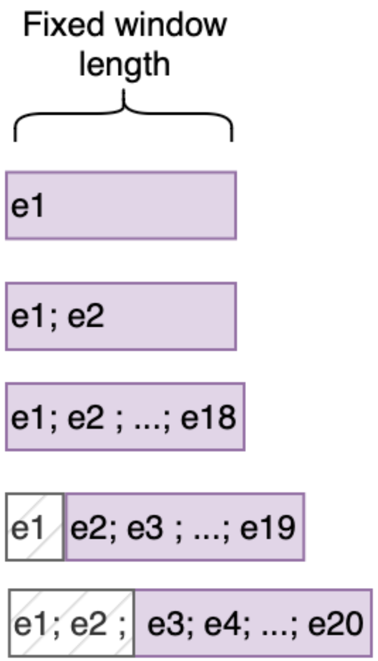
Second approach: Seq2Seq-DDP + Transition



State	Top EDU in Buffer	Action
$c_s = (e_0, \epsilon)$	e_1	$(e_1, e_0), \text{root}$
$c_1 = (e_1, \{(e_1, e_0, \text{root})\})$	e_2	$(e_2, e_1), \text{elaboration}$
$c_2 = (e_2, \{(e_1, e_0, \text{root}), (e_2, e_1, \text{elaboration})\})$	e_3	$(e_3, e_1), \text{qa-pair}$
...	...	...
$c_n = (e_n, \{(e_1, e_0, \text{root}), (e_2, e_1, \text{elaboration}), \ldots, (e_n, e_{n-1}, \text{qa-pair})\})$	$\emptyset$	$\emptyset$

Second approach: Seq2Seq-DDP + Transition

- Y(natural) is a sequence of elements with a structure: e_i is r_{ki} of e_k
- Y(focused) is a sequence of elements with structure: $^{**}e_i|r_{ki}$ of e_k
- Sliding window strategy to cope with increasing input length:
  - The closest EDUs are the most relevant to the target EDU
- Last step: from Y(natural) and Y(focused) to the target discourse graph is easy! No worry of mismatched EDUs or *counting* issue.



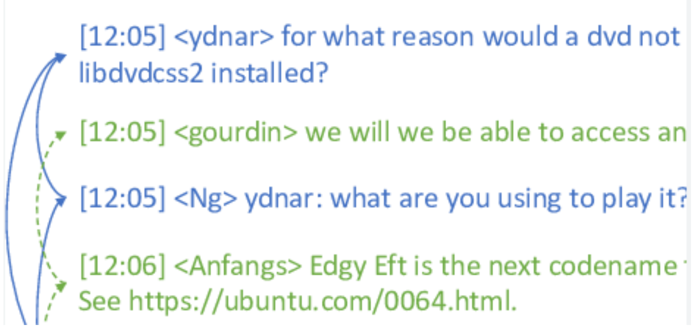
Evaluation: Datasets

- Test on two dialogue datasets
  - **STAC** (The Settlers of Catan game): 1,000 gaming conversations, ~10k discourse units
  - **Molweni** (Ubuntu Chat logs): 10,000 short log conversations, ~80k discourse units

	Train			Development			Test		
Dataset	#Doc	#Sent	#Token	#Doc	#Sent	#Token	#Doc	#Sent	#Token
STAC	911	$10k$	$47k$	97	$1k$	$5k$	109	$1k$	$5k$
Molweni	9000	$79k$	$945k$	500	$4k$	$52k$	500	$4k$	$52k$

- Metric: micro-F1 score
- **T0-3B** checkpoint as backbone model


STAC



Molwnei

Evaluation: Simple Seq2Seq-DDP

		STAC				Molweni			
	System	Link	Δ	Full	Δ	Link	Δ	Full	Δ
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1	
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0	

Overall, fine-tuned T0-3B model can perform well on discourse parsing

- On Molweni, Y(natural) and Y(augmented) both perform well

Evaluation: Simple Seq2Seq-DDP

		STAC				Molweni			
	System	Link	Δ	Full	Δ	Link	Δ	Full	Δ
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1	
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0	

Overall, fine-tuned T0-3B model can perform well on discourse parsing

- On STAC, more pronounced difference between Y(natural) and Y(augmented)
  - STAC contains shorter EDUs, similar ones occur
  - Y(natural) omits the text and only use EDU markers → cause ambiguity

Evaluation: Simple Seq2Seq-DDP

		STAC				Molweni				STAC		Molweni	
	System	Link	Δ	Full	Δ	Link	Δ	Full	Δ	Hallu	Miss	Hallu	Miss
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1		3.1%	1.7%	0.4%	0
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0		0	0.2%	0	0

Problems

- Hallucinated EDUs
- Missed EDUs
- Incorrect *counting*

y	$\hat{y}$
y_{nat}: …[e_{14}] is Acknowledgement of [e_{13}] ; [e_{15}] is Continuation of [e_{13}] ; [e_{16}] is Elaboration of [e_{15}].	$\hat{y}_{nat}$: [e_{14}] is Acknowledgement of [e_{12}] ; [e_{15}] is Result of [e_{14}] ; [e_{16}] is QA-pair of [e_{15}] ; [e_{17}] is Contrast of [e_{16}].
y_{nat}: [e_0] is root; [e_1] is Acknowledgement of [e_0] ; [e_2] is Elaboration of [e_1] ; …[e_{29}] is Clarification_question of [e_{28}] ; [e_{30}] is Correction of [e_{29}] ; [e_{31}] is Clarification_question of [e_{28}] ; [e_{32}] is QA-pair of [e_{29}] ; [e_{33}] is Explanation of [e_{32}] ; [e_{34}] is QA-pair of [e_{31}] ; [e_{35}] is Comment of [e_{32}] ; [e_{36}] is Comment of [e_{32}].	$\hat{y}_{nat}$: [e_0] is root; [e_1] is Acknowledgement of [e_0] ; [e_2] is Continuation of [e_1] ; …[e_{29}] is Comment of [e_{28}] ; [e_{30}] is Comment of [e_{28}] ; [e_{30}] is Comment of [e_{28}] ; [e_{30}] is Comment of [e_{28}] ; [e_{30}] is Comment of [e_{28}]

Evaluation: Simple Seq2Seq-DDP

		STAC				Molweni				STAC		Molweni	
	System	Link	Δ	Full	Δ	Link	Δ	Full	Δ	Hallu	Miss	Hallu	Miss
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1		3.1%	1.7%	0.4%	0
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0		0	0.2%	0	0

Problems

- Hallucinated EDUs
- Missed EDUs
- Incorrect *counting*

y_{aug}: [ztime: morning | e_1 | root = e_0] [Shawnus: hey | e_1 | Acknowledgement = e_0] [Shawnus: good morning | e_2 | Elaboration = e_1] ...[Shawnus: misplaced/ | e_{29} | Clarification_question = e_{28}] [Shawnus: ? | e_{30} | Correction = e_{29}] [somdechn: Need to undo are you? | e_{31} | Clarification_question = e_{28}] [ztime: no. | e_{32} | QA-pair = e_{29}] [ztime: you took the spot I was looking at. | e_{33} | Explanation = e_{32}] [ztime: no it's fine. | e_{34} | QA-pair = e_{31}] [Shawnus: hahaha | e_{35} | Comment = e_{32}] [somdechn: Got to be mean here. | e_{36} | Comment = e_{32}]

$\hat{y}_{aug}$: [ztime: morning | e_1 | root = e_0] [Shawnus: hey | e_1 | Acknowledgement = e_0] [Shawnus: good morning | e_2 | Continuation = e_1] ...[Shawnus: misplaced/ | e_{25} | QA-pair = e_{24}] [Shawnus:? | e_{25} | Continuation = e_{24}] [somdechn: Need to undo are you? | e_{25} | Clarification_question = e_{24}] [ztime: no. | e_{25} | QA-pair = e_{24}] [ztime: you took the spot I was looking at. | e_{25} | Explanation = e_{24}] [ztime: no it's fine. | e_{25} | Acknowledgement = e_{24}] [Shawnus: hahaha | e_{25} | Comment = e_{24}] [Shawnus: hahaha | e_{27} | Comment = e_{24}] [Shawnus: hahaha | e_{27}

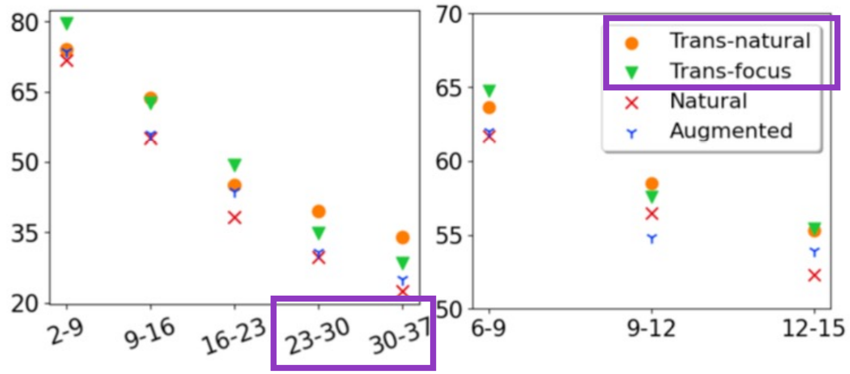
placed/ | e_{25} | QA-pair = e_{24}] [Shawnus:? | e_{25} | Continuation = e_{24}] [somdechn: Need to undo are you? | e_{25} | Clarification_question = e_{24}] [ztime: no. | e_{25} | QA-pair = e_{24}] [ztime: you took the spot I was looking at. | e_{25} | Explanation = e_{24}] [ztime: no it's fine. | e_{25} | Acknowledgement = e_{24}] [Shawnus: hahaha | e_{25} | Comment = e_{24}] [Shawnus:

Evaluation: Seq2Seq-DDP+Transition

		STAC				Molweni				STAC		Molweni	
	System	Link	Δ	Full	Δ	Link	Δ	Full	Δ	Hallu	Miss	Hallu	Miss
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1		3.1%	1.7%	0.4%	0
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0		0	0.2%	0	0
Y (natural)	Seq2Seq-DDP+Transition	70.8 ± 0.9	↑ 5.2	55.1 ± 1.0	↑ 8.2	83.5 ± 0.2	↑ 2.1	60.3 ± 0.1	↑ 2.5	-	-	-	-
Y (focused)	Seq2Seq-DDP+Transition	72.3 ± 0.6	↑ 5.5	56.6 ± 0.6	↑ 4.6	83.4 ± 0.6	↑ 1.0	60.0 ± 0.5	↑ 0.9	-	-	-	-

Comparison

- Ours: Seq2Seq-DDP+Transition largely **outperforms its counterpart**, with superior performance on longer documents



X axis: #EDUs in a doc. Y axis: F1 score

Evaluation: Seq2Seq-DDP+Transition

	System	STAC Link	Δ	STAC Full	Δ	Molweni Link	Δ	Molweni Full	Δ	STAC Hallu	STAC Miss	Molweni Hallu	Molweni Miss
Y (natural)	Seq2Seq-DDP	65.6 ± 0.3		46.9 ± 1.8		81.4 ± 0.4		57.8 ± 0.1		3.1%	1.7%	0.4%	0
Y (augmented)	Seq2Seq-DDP	66.7 ± 0.7		52.0 ± 0.1		82.4 ± 0.4		59.1 ± 1.0		0	0.2%	0	0
Y (natural)	Seq2Seq-DDP+Transition	70.8 ± 0.9	↑ 5.2	55.1 ± 1.0	↑ 8.2	83.5 ± 0.2	↑ 2.1	60.3 ± 0.1	↑ 2.5	-	-	-	-
Y (focused)	Seq2Seq-DDP+Transition	72.3 ± 0.6	↑ 5.5	56.6 ± 0.6	↑ 4.6	83.4 ± 0.6	↑ 1.0	60.0 ± 0.5	↑ 0.9	-	-	-	-
Shi and Huang (2019)	GRU+Pointer*	72.9 ± 0.4		54.2 ± 0.5		77.9 ± 0.4		54.1 ± 0.6					
Liu and Chen (2021)	RoBERTa+Pointer	72.9 ± 1.5		57.0 ± 1.0		79.0 ± 0.4		55.4 ± 1.8					
Chi and Rudnicky (2022)	RoBERTa+CLE†	73.0 ± 0.5		58.1 ± 0.7		81.0 ± 0.7		58.6 ± 0.6					
Li et al. (2023c)	BERT+Biaffine+Pointer	73.0		58.5		83.2		59.8					

Comparison

- SOTA models: comparable results with our Seq2Seq+Transition models
  - Ours do not need specific parsing modules or modification of LLM architecture
  - Ours can predict richer graph-like structures thanks to flexible Y scheme e_i is r_{ki} of e_k r_{mi} of e_m r_{ni} of e_n

Further Investigation: Label Semantics & Abridged Output

- Y (natural): e_i is elaboration of e_j
- Y(masked): e_i is rel4 of e_j

- Y (natural): e_i is elaboration of e_j
- Y(abridged): $e_i e_j$ rel4

- On STAC:
  Link prediction -2%
  Link+Relation prediction -9%

- On Molweni:
  No significant performance drop

Further Investigation: Label Semantics & Abridged Output

- Y (natural): e_i is elaboration of e_j
- Y(masked): e_i is rel4 of e_j

- Y (natural): e_i is elaboration of e_j
- Y(abridged): $e_i e_j$ rel4

- STAC 900 train docs *vs.* Molweni 9,000 train docs

- On STAC:
  Link prediction -2%
  Link+Relation prediction -9%

- On Molweni:
  No significant performance drop

→ Label Semantics and *natural language–like scheme* brings more accurate predictions, especially when training data is of low volume

→ Sufficient Supervision enables us to use the simpler format

Further Investigation: Pretrained LLMs

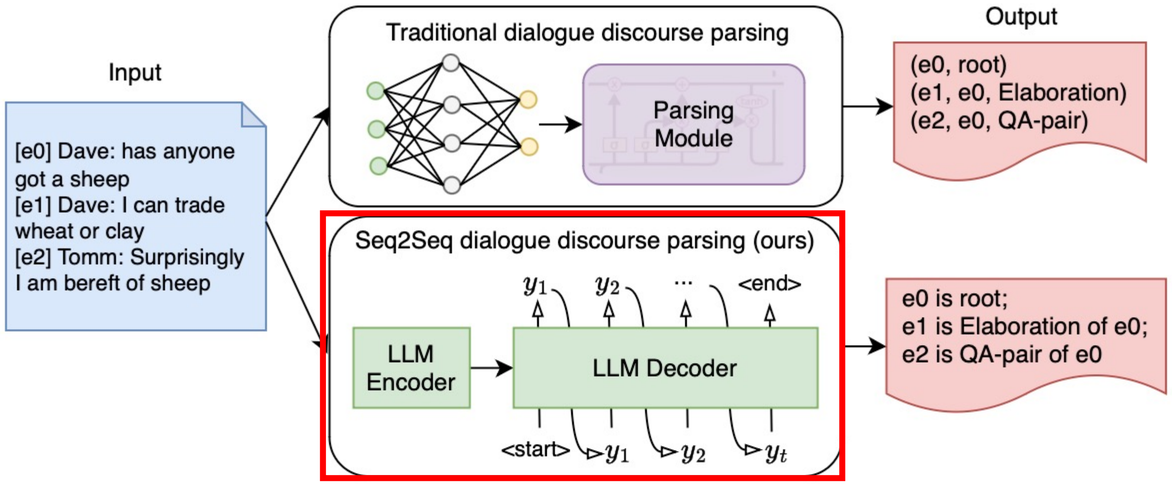
- Models: T5, Flan-T5, T0
- Sizes: 250M, 780M, 3B

Pre-trained model	#Params	Link (F_1)	Full (F_1)
T5-large	738M	59.3 ± 0.6	36.4 ± 0.6
T5-3B	3B	60.7 ± 1.3	40.5 ± 0.9
Flan-T5-base	250M	63.0 ± 0.5	36.7 ± 0.1
Flan-T5-large	780M	67.2 ± 1.4	46.6 ± 1.8
Flan-T5-xl	3B	$\underline{68.5} \pm 0.5$	$\mathbf{50.4} \pm 0.1$
T0-3B	3B	$\mathbf{69.2} \pm 0.5$	$\underline{50.2} \pm 0.7$

- Flan-T5 and T0 comparable results
- Both largely exceed T5 (up to 2-digit gains)

→ Instruction tuning enhances model's ability in learning complex reasoning task.

Summary and Perspectives



Our Research Goal:

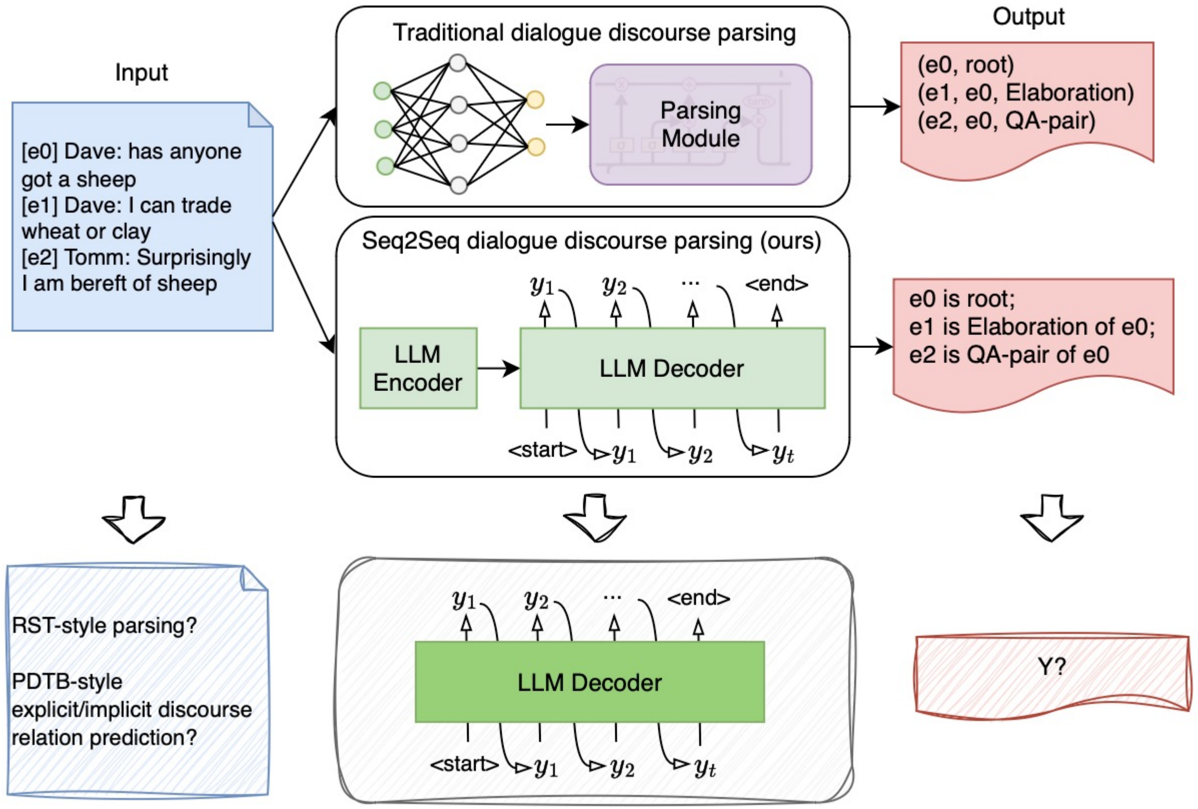
Leverage LLMs for discourse structure prediction without explicitly designing parsing modules or changing the architecture of LLMs.

This Study:

Turn parsing task into a seq2seq generation task;

Propose two seq2seq-DDP approaches with sophisticated output schemes

Summary and Perspectives



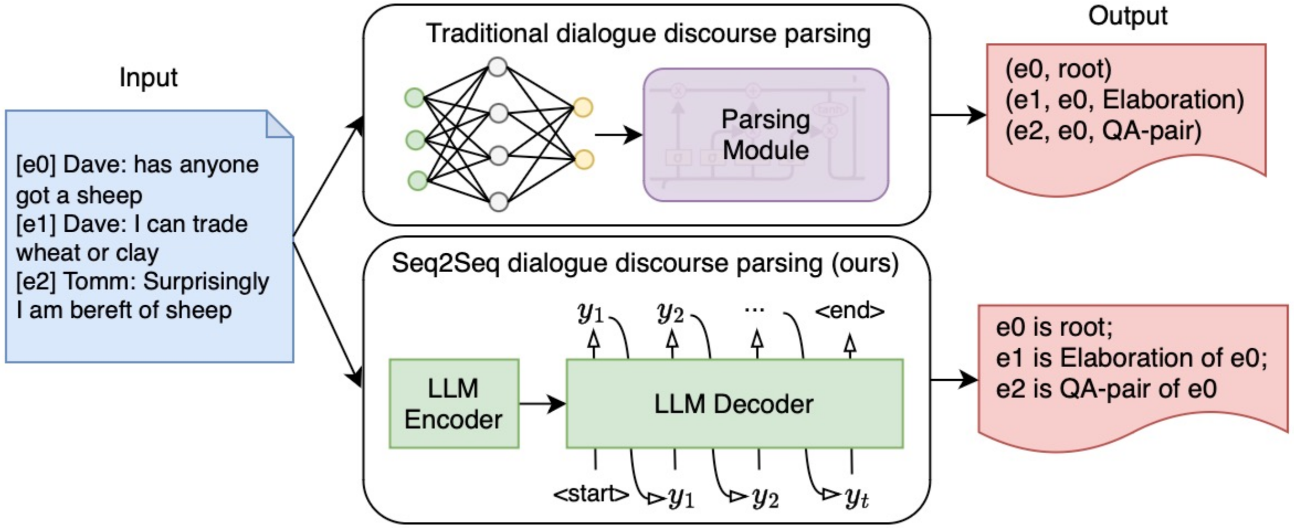
Future Directions:

Extend our method to other discourse parsing tasks: e.g., RST, PDTB, which may require alternative sequence representations.

- RST-style parsing with Llama2 [Maekawa et al., 2024], EACL

Explore generative open-source model architectures.

Dialogue Discourse Parsing as Generation: a Seq-to-Seq LLM-based Approach



Thank you!

Chuyuan Li, Yuwei Yin, Giuseppe Carenini

University of British Columbia

SIGdial 2024, September 7,

Kyoto University

References

- [Zhang et al., 2023] Wenzheng Zhang, Sam Wiseman, and Karl Stratos. 2023. Seq2seq is all you need for coreference resolution. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 11493–11504, Singapore. Association for Computational Linguistics.
- [Bohnet et al., 2023] Bernd Bohnet, Chris Alberti, and Michael Collins. 2023. Coreference resolution through a seq2seq transition- based system. Transactions of the Association for Computational Linguistics, 11:212–226.
- [Paolini et al., 2021] Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cícero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages. In 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net.
- [Rongali et al., 2020] Subendhu Rongali, Luca Soldaini, Emilio Monti, and Wael Hamza. 2020. Don't parse, generate! A sequence to sequence architecture for task-oriented se- mantic parsing. In WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020, pages 2962– 2968. ACM / IW3C2.
- [He and Choi, 2023] Han He and Jinho D Choi. 2023. Unleashing the true potential of sequence-to-sequence models for sequence tagging and structure parsing. Transactions of the Association for Computational Linguistics, 11:582– 599.
- [Kojima et al., 2022] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yu- taka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. Advances in neural information processing systems, 35:22199– 22213.
- [Nivre 2023] Joakim Nivre. 2003. An efficient algorithm for projective dependency parsing. In Proceedings of the Eighth International Conference on Parsing Technologies, pages 149–160, Nancy,
- [Nivre 2028] Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. Computational Linguistics, 34(4):513–553.
- [Maekawa et al., 2024] Aru Maekawa, Tsutomu Hirao, Hidetaka Kamigaito, and Manabu Okumura. 2024. Can we obtain significant success in RST discourse parsing by using large language models? In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2803–2815, St. Julian's, Malta. Association for Computational Linguistics.
- [Thompson et al., 2024] Thompson, Kate, Akshay Chaturvedi, Julie Hunter, and Nicholas Asher. "LLaMIPa: An Incremental Discourse Parser." arXiv preprint arXiv:2406.18256 (2024).

The age of Large Generative Models

Method	STAC Link	STAC Link&Rel
Afantenos et al. (2015)	68.8	50.4
Perret et al. (2016)	68.6	52.1
Shi and Huang (2019)	73.2	55.7
ChatGPT zero w/ desc.	20.5	4.3
ChatGPT zero w/o desc.	20.0	4.4
ChatGPT few (n=1) w/ desc.	21.0	7.1
ChatGPT few (n=3) w/ desc.	20.7	7.3
ChatGPT few (n=1) w/o desc.	21.2	6.2
ChatGPT few (n=3) w/o desc.	21.3	7.4

Bad results in directly prompting T0 on discourse parsing.

Similarly, GPT-3.5 on dialogue discourse parsing [Chan et al., 2023]

- Zero-shot and few-show In-context learning
- With and without label description
- Only to find abysmal results

Existing work on dialogue discourse parsing

Model	STAC	
	Link	Link&Rel
MST (Afantenos et al., 2015)	68.8	50.4
ILP (Perret et al., 2016)	68.9	53.1
Deep Sequential (Shi and Huang, 2019)	73.2	55.7
Struct-Aware GNN (Wang et al., 2021a)	73.5	57.3
Hierarchical Transformer-based (Liu and Chen, 2021)	**75.3**‖	56.9‖
QA-DP Multi-task (He et al., 2021)	-	-
DiscProReco Multi-task (Yang et al., 2021)	74.1*	57.0*
Distance-Aware Multi-task (DAMT) (Fan et al., 2022)	73.6	57.4
SSP+SCIJE (Yu et al., 2022)	73.0	57.4
Struct-Joint (Chi and Rudnicky, 2022)	74.4	**59.6**

Various decoding strategy

- Maximum spanning tree decoders [Muller et al., 2012]
- Integer linear programming [Perret et al., 2016]

Neural models

- Deep sequential + classification [Shi and Huang, 2019]
- Pre-trained language model (PLM) + classification [Liu and Chen, 2021]
- Graph neural network [Wang et al., 2021]

However, ...

- Heavy feature engineering, specialized decoding strategies
- Mostly limited to *trees*
- No use of latent knowledge in recent Large Generative Models

First system: Seq2Seq-DDP