The Role of Semantic Parsing in Understanding Procedural Text

Hossein Rajaby Faghihi, Parisa Kordjamshidi, Choh Man Teng, James Allen

Michigan State University
Florida Institute for Human and Machine Cognition

01

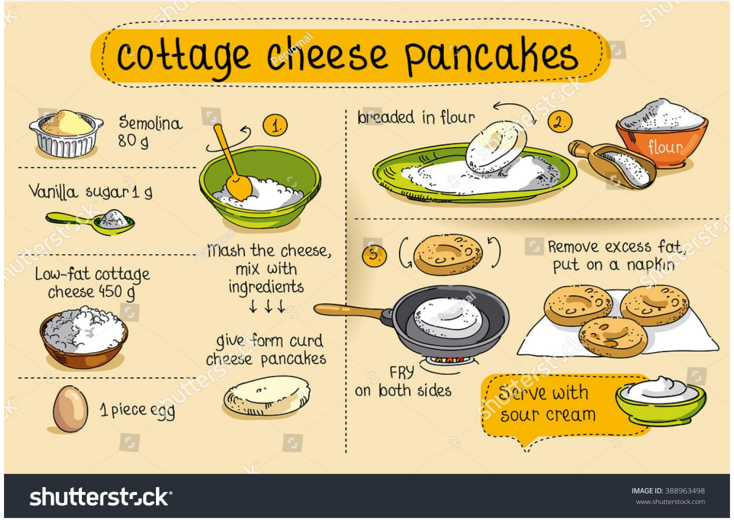
Problem & Task

What do we do?!

2023

2023

Procedural Text



Process

Describes a series of action which changes the status of the world from one point to another

Steps

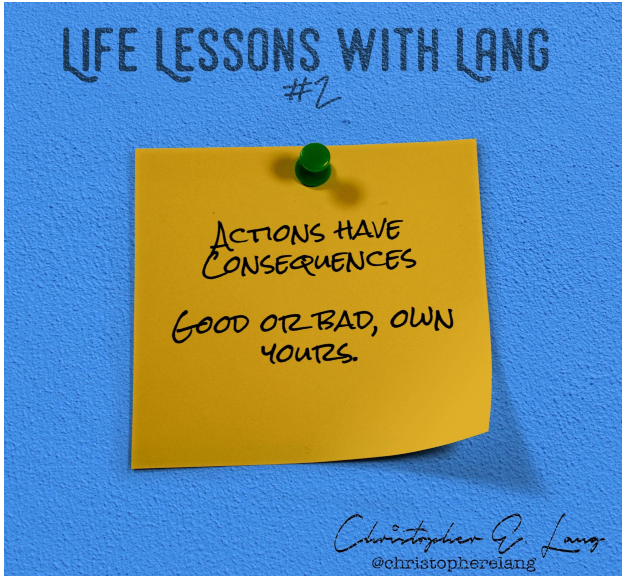
Process can be divided into steps. Each step would contain some actions and some consequences.

World

The world can be defined by a set of entities and their status. Through the process, the existence of these entities and their state would change.

2023

Procedural Reasoning



Procedural Reasoning

Is to reason about the process in terms of its effects on the world.

Action

An action is a change to the world. It can be limited to a number of entities.

Consequence

An action can have multiple consequences. The consequence of one action can be different on various entities involved in the action

Benchmark

		Participants				
Paragraph	State number	Water	Light	CO2	Mixture	Sugar
(Before the process starts)	State 0	Soil	Sun	?	-	-
Roots absorb water from soil	State 1	Root	Sun	?	-	-
The water flows to the leaf	State 2	Leaf	Sun	?	-	-
Light from the sun and CO2 enter the leaf	State 3	Leaf	Leaf	Leaf	-	-
The water, light, and CO2 combine into a mixture	State 4	-	-	-	Leaf	-
Mixture forms sugar	State 5	-	-	-	-	Leaf

Table 1: An example of procedural text and its annotations from the Propara dataset (Dalvi et al., 2018). "-" means entity does not exist. "?" means the location of entity is unknown.

02

Intuition

What drives us?!

Reasoning over Both Local and Global Context

(context) Step 1: **Magma** rises to the surface, Step 2: **Magma** cools down to form **lava**.

(Semantic Frame)

(Predicate: Form, Affected: Magma, Result: Lava)

(Chain of reasoning)

1. Location of **Magma** at **step 1**: **Surface**
2. (SRL) **Magma** is the **consumed** in making **Lava**
3. (Common-sense) In conversion, the location of the **result** matches the location of the consumed (**affected**) object.
4. The location of **Lava** is **Surface**.

Common-sense through ontological features

(context) Step 1: **Generator** makes **electricity**.

(Semantic Frame)

(Predicate: makes, Agent: Generator, Result: Electricity)

(Chain of reasoning)

1. (Common-sense) **Agent** and **result** are co-located
2. (Common-sense) The **result** can be **inside** the **Agent,** if agent is a container.
3. (Ontology) **Generator** is a **container**
4. The location of **Electricity** is **Generator**.

03

Semantic Parsing & Representation

How is the knowledge extracted and represented?!

Semantic Parsing Source

- We use two sources for semantic parsing of the sentences
  - A shallow parser, that is a deep neural model trained with data, for **semantic role labeling (SRL)**
  - A symbolic deep semantic parser **(TRIPS)**

Semantic Parsing Graphs

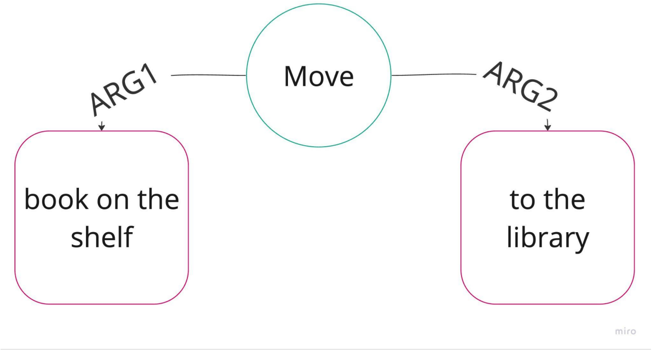
- The semantic relationships in each sentence is depicted as a graph.
- The relations between each sentence to another relies on exact-match of entities and candidate locations and their co-references.

2023

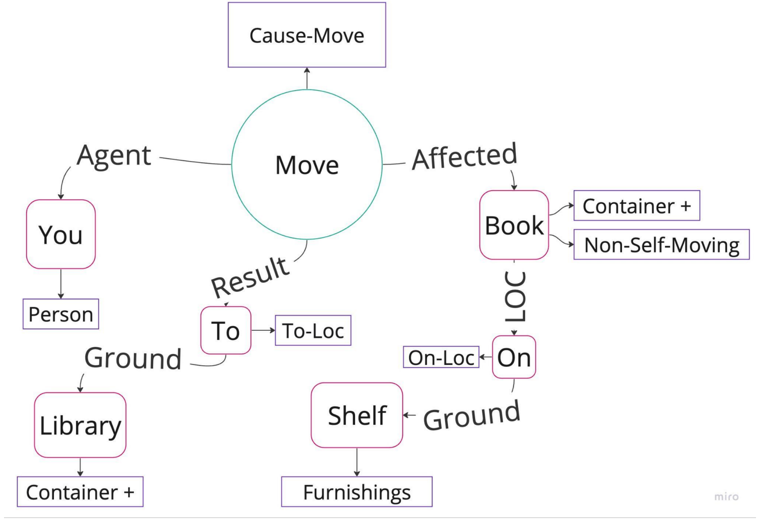
Semantic Parsing Graphs

Sentence: Move the book on the shelf to the library

SRL graph



TRIPS graph



04

Symbolic Procedural Reasoning

Can we use semantic parsing as a standalone model for procedural reasoning?

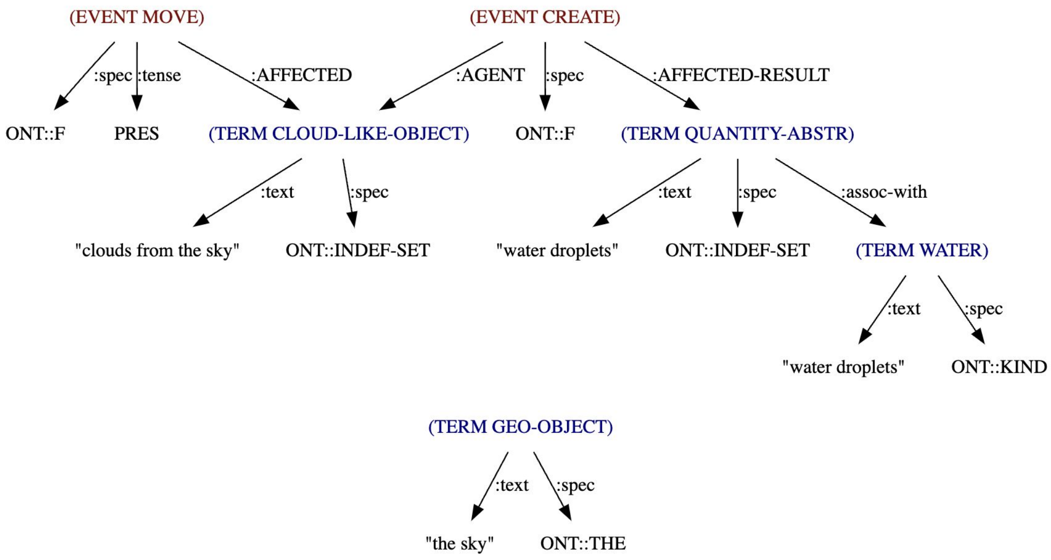
PROPOLIS: Symbolic Procedural Reasoning

Propolis uses the Trips parses to directly predict the actions and locations of entities in procedural texts. It consists of three modules:

- **Graph Abstraction**:
  - Maps the graph to the set of actions Move, Destroy, or Create
  - Selects the important roles for each predicate: agent, affected, result, ...
  - Generalization over different verbs:
    - Put, Move, place → Move
    - Form, Create → Create

PROPOLIS: Symbolic Procedural Reasoning

- **Graph Abstraction**:
  - Sentence: clouds from the sky collide in order to create water droplets.



2023

PROPOLIS: Symbolic Procedural Reasoning

➔ **Rule-based Local Decisions:**

- Uses the abstracted graph to make local decisions about entities
- Which role of the predicate is the one being created, which is being destroyed and which is being moved
- What is the location of the moved or created object in the semantic frame

Main Predicate	Roles	Decisions
Move	Affected, Agent	The "Affected" is being moved.
Move	Agent	The "Agent" is being moved.
Destroy	Affected	The "Affected" is being destroyed
Create	Affected_Result, Affected	The "Affected_Result" is being created
Create	Affected	The "Affected" is being created
Change	Affected, Res	The "Affected" is being destroyed, and the "Res" is being created

Table 1: The rules used to evaluate the effect of actions on various roles of the semantic frame

PROPOLIS: Symbolic Procedural Reasoning

- **Global Reasoning:**
  - Make sure that the set of local decisions are correct!
  - Action consistency
    - Initial prediction: Move, Create, Exist
    - After applying consistency rules: Move, Exist, Exist
  - Location and Action Consistency and hidden actions
    - Initial prediction: None (loc 1, loc 2)
    - After applying consistency rules: Move (loc 1, loc 2) or None (loc 1, loc 1)

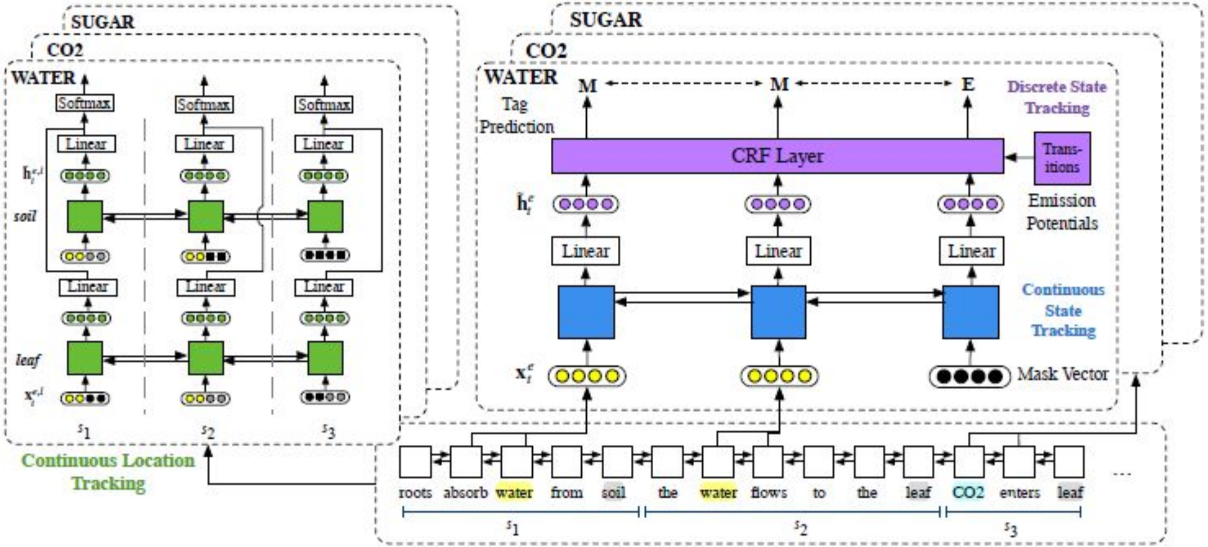
05

Neuro-Symbolic Reasoning

How to integrate semantic information with neural models?!

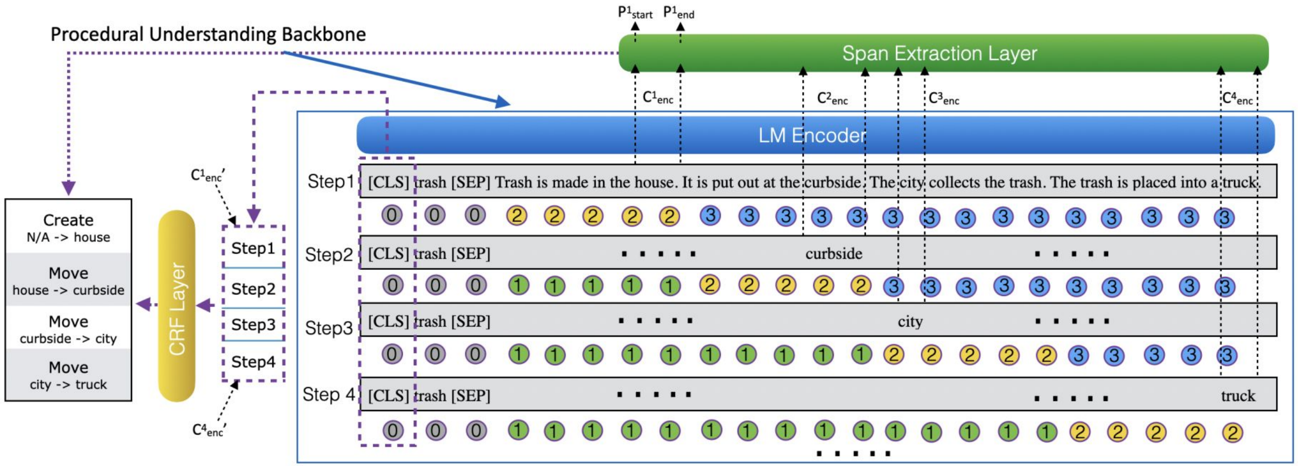
Neural Backbones

- We choose two neural backbones for the integration:
  - NCET



Neural Backbones

- We choose two neural backbones for the integration:
  - NCET
  - TSLM (CGLI)

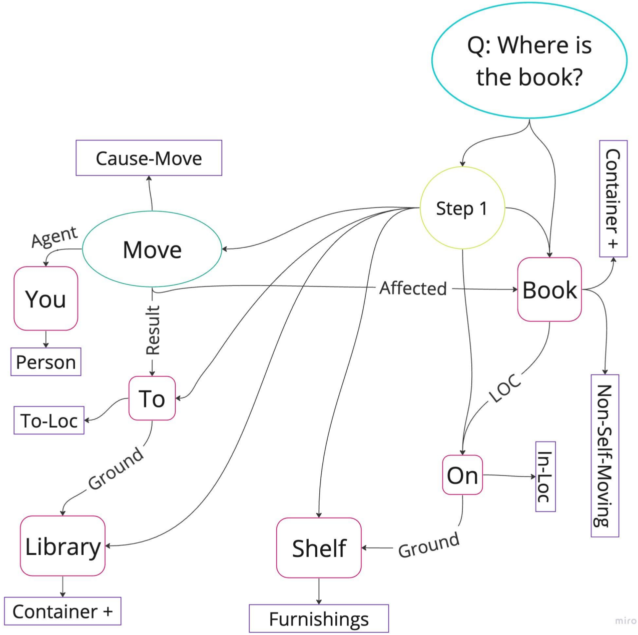


Integration

- Using Graph Attention encoder (GAT) to encode the semantic graphs.

- New token representations:
    - Neural Backbone + GAT

- For CGLI model, the graph is updated with the question:



06

Evaluation & Experiments

How do the models perform?!

2023

High-level evaluations

- Sentence Level
  - Cat1: evaluates whether an entity “e” has been created (destroyed/moved) during the process
  - Cat2: evaluates when an entity “e” is created (destroyed/moved).
  - Cat3: evaluates where “e” is created (destroyed/moved)
- Document Level
  - Inputs
  - Outputs
  - Conversion
  - Moves

Model's Performance

#Row	Models	Sentence-level evaluation					Document-level evaluation		
		Cat1	Cat2	Cat3	Macro-avg	Micro-avg	Precision	Recall	F1
1	ProLocal	62.7	30.5	10.4	34.5	34.0	**77.4**	22.9	35.3
2	ProGlobal	63	36.4	35.9	45.1	45.4	46.7	52.9	49.4
3	KG-MRC	62.9	40	38.2	47	46.6	64.5	50.7	56.8
4	PROPOLIS(ours)	69.9	37.71	5.6	37.74	36.67	70.9	50.0	58.7
5	NCET (re-implemented)	75.54	45.46	41.6	54.2	54.38	68.4	63.6	66
6	REAL(re-implemented)*	78.9	48.31	41.62	56.29	56.35	67.3	64.9	66.1
7	NCET + SRL(ours)	77.1	46.35	42	55.16	55.32	67.8	65.2	66.5
8	NCET + TRIPS(ours)	77.1	48.12	43.36	56.19	56.32	72.5	65.4	68.8
9	NCET + TRIPS(Edge)(ours)	75.68	47.6	45.71	56.33	56.37	69.9	65.5	67.6
10	NCET + PROPOLIS(ours)$^{+}$	78.54	48.69	44.26	57.16	57.31	74.6	65.8	69.9
11	DynaPro	72.4	49.3	44.5	55.4	55.5	75.2	58	65.5
12	KOALA	78.5	53.3	41.3	57.7	57.5	77.7	64.4	70.4
13	TSLM	78.81	56.8	40.9	58.83	58.37	68.4	68.9	68.6
14	CGLI	80.3	**60.5**	48.3	**63.0**	**62.7**	74.9	**70**	**72.4**
15	CGLI + TRIPS (ours)	**80.62**	58.94	**49.08**	62.88	62.68	74.5	68.5	71.4

Decision-Level Evaluation

- The prior metrics do not directly evaluate the model's decisions
- Our new metric directly evaluate the models decision. The decisions are categorized based on their difficulty:
- Categories of decisions: Exists, Does not Exist
  - **Local Decision**: ✔ (Location, Local sentence) ✔(Entity, Local Sentence)
  - **Global Location**: ✖ (Location, Local Sentence) ✔ (Entity, Local Sentence)
  - **Global Entity**: ✖ (Entity, Local Sentence) ✔ (Location, Local Sentence)
  - **Global Decision**: ✖ (Location, Local Sentence) ✖ (Entity, Local Sentence)
  - **Locally Ambiguous**: A local decision where, Multiple verb and actions

	Local	Global Loc	Global Ent		Global Loc and Ent	Ambiguous
	Both	Both	Actions	Locations	Both	Actions
Train	885	367	438	340	114	593
Dev	116	44	66	3	9	76
Tests	105	61	98	71	18	110

Model's Performance

Model	Local			Global Loc			Global Ent			Global Loc and Ent			Amb+
	A	L	Both	A	L	Both	A	L	Both	A	L	Both	A
KOALA	74.3	65.7	59.0	**86.9**	24.6	22.9	1.0	7.0	0.0	5.6	11.1	0	73.63
PROPOLIS	55.2	19.0	19.0	63.9	1.6	1.6	0.0	9.9	0.0	0.0	0.0	0.0	52.7
NCET	69.5	62.8	60.0	70.5	36.1	29.5	3.1	5.6	0.0	0.0	0.0	0.0	57.2
NCET + SRL	68.6	65.7	61.9	77.0	36.1	31.1	10.2	5.6	0.0	5.5	5.5	0.0	62.7
NCET + TRIPS	71.4	67.6	**63.8**	75.4	42.6	36.1	10.2	9.9	2.8	5.5	11.1	0.0	63.6
NCET + PROPOLIS	71.4	64.8	61.9	83.6	36.1	34.4	3.1	7.0	0.0	5.5	5.5	0.0	70.9
CGLI	65.7	62.9	54.3	75.4	59.0	50.8	**19.4**	19.7	11.3	22.2	**27.8**	11.1	70.0
CGLI + TRIPS	**75.2**	**70.5**	61.9	80.3	**60.6**	**52.2**	17.3	**22.5**	**12.7**	**27.8**	**27.8**	**16.7**	**74.5**

20
23

07

TakeAway

What do we conclude?

Conclusion

- We Proposed new evaluation metrics for a procedural reasoning benchmark, Propara:
    - Based on the difficulty of the low-level decisions.
    - Hardest decision is where the entity and location do not appear in the local context. Decisions over global entity is harder than global location.

- Symbolic Procedural reasoning → outperforms neural models; Especially when pre-trained language models are not used!

- Integration with SOTA models → helpful in most of the metrics.

- Integration → complimentary effect on the original backbone.
    - If the baseline performed better on local decision, the improvement is higher on global decisions and vice-versa.

Thank you!

rajabyfa@msu.edu