Semantic-Aware Logical Reasoning via a Semiotic Framework

Yunyao Zhang[1], Xinglang Zhang[1], Junxi Sheng[1], Wenbing Li[1], Junqing Yu[1], Yi-Ping Phoebe Chen[2], Zikai Song[1]†
[1] Huazhong University of Science and Technology, Wuhan, China
[2] La Trobe University, Melbourne, Australia

ACL 2026

1 Motivation

Why Logical Reasoning Needs Semantics?

Wittgenstein

The meaning of language depends on use, context and form of life.

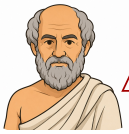
Meaning is Contextual

Plato's *Republic*

What is justice? Is repaying debt always just?

Meaning is debated

Aristotle and Logic



To know the good, We must know what is not.

Meaning needs relations

★Key Question

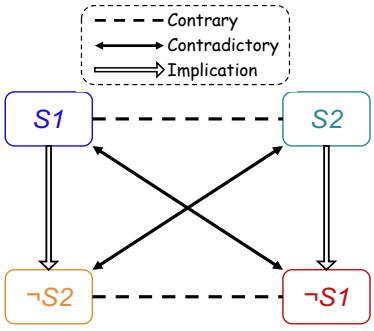
How can we reason logically when meaning itself is ambiguous?

2 Semiotic Square

Greimas' Semiotic Square

Example(Justice)

S1 NL: Repayment of debt is always just. FOL: $\forall x\,(Debt(x) \wedge Repaid(x) \rightarrow Just(x))$	**S2** NL: Repayment of debt is always unjust. FOL: $\forall x\,(Debt(x) \wedge Repaid(x) \rightarrow \neg Just(x))$
¬S1 NL: Some repaid debts are unjust. FOL: $\exists x\,(Debt(x) \wedge Repaid(x) \wedge \neg Just(x))$	**¬S2** NL: Some repaid debts are just. FOL: $\exists x\,(Debt(x) \wedge Repaid(x) \wedge Just(x))$



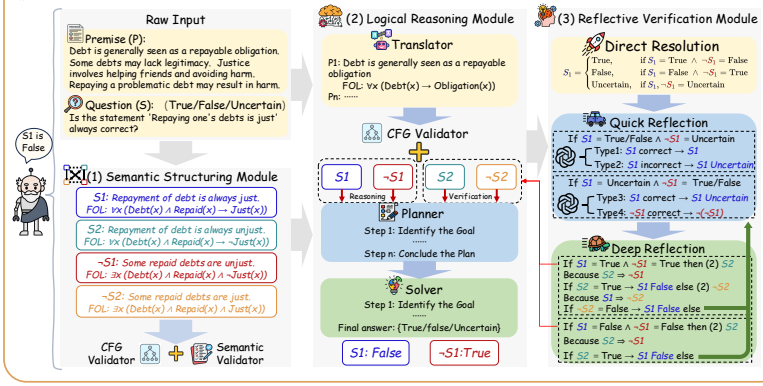
By leveraging Greimas' Semiotic Square, LLM can turn hidden semantics into explicit logical structure.

Unified rules for constructing contraries and contradictories

#	S_1	$\neg S_1$	S_2	Constraint
1	$\forall x\,\varphi(x)$	$\exists x\,\neg\varphi(x)$	$\forall x\,\neg\varphi(x)$	N/A
2	$A \wedge B$	$\neg A \vee \neg B$	$A \wedge \neg B$	N/A
3	$A \leftrightarrow B$	$A \oplus B$	$A \leftrightarrow \neg B$	N/A
4	$\exists x\,\varphi(x)$	$\forall x\,\neg\varphi(x)$	$\exists x\,\neg\varphi(x)$	$D = \emptyset$
5	$A \rightarrow B$	$A \wedge \neg B$	$A \rightarrow \neg B$	Sat(A)
6	$A \vee B$	$\neg A \wedge \neg B$	$A \vee \neg B$	$A = \mathbf{F}$

3 Method & RepublicQA

LogicAgent: Three-stage Semantic-aware reasoning



RepublicQA: A benchmark for semantic-aware logical reasoning

Basics

- 600 examples
- Abstract philosophy : Classical Philosophy Books
- True | False | Uncertain
- Contrary & Contradictory forms

Semantic complexity Comparison

Benchmark	Total	FKGL↑ (Concept.)	TTR↑ (Lexical Diversity)	MTLD↑ (Lexical Diversity)	UBR↑ (Lexical Diversity)	Contr.↑ (Struct.)
ProntoQA	500	6.78	0.448	13.93	0.852	0.00
FOLIO	204	6.62	0.569	33.54	0.805	0.30
ProofWriter	600	1.25	0.193	11.31	0.513	0.00
ProverQA	500	8.44	0.616	34.84	0.774	0.13
RepublicQA	600	**11.94**	**0.685**	**74.81**	**0.929**	**0.70**
		+41.5%	+11.2%	+114.7%	+9.0%	+133.3%

4 Experiments

A. Main Results

Table 3: Performance comparison across RepublicQA and other logical reasoning benchmarks. Best results are in **bold**, second-best are underlined.

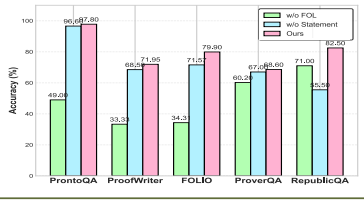
Type	Method	RepublicQA Qwen2.5↑	RepublicQA GPT-4o↑	RepublicQA Avg↑	Other Benchmarks Pronto↑	ProofWriter↑	FOLIO↑	ProverQA↑	Avg↑
LR	Naive	68.50	74.00	71.25	82.00	59.17	60.29	39.60	60.27
	CoT	72.00	75.00	73.50	92.40	63.17	68.42	47.20	67.80
AR	CR	57.00	71.00	64.00	80.20	58.33	71.57	51.80	65.48
	ToT	56.00	69.50	62.75	82.50	64.40	72.54	53.40	68.21
SR	Logic-LM	70.00	73.50	71.75	91.89	63.82	71.93	62.40	72.51
	SymbCoT	76.00	80.50	78.25	95.20	64.67	70.59	57.20	71.92
AR+SR	Aristotle	74.50	82.50	78.50	94.80	63.23	68.68	56.20	70.73
	LogicAgent	**82.50** (+6.50)	**87.00** (+4.50)	**84.75** (+6.25)	**97.80** (+2.60)	**71.95** (+7.28)	**79.90** (+7.97)	**68.60** (+6.20)	**79.56** (+7.05)

Finding 1: Semantic Structuring improves both RepublicQA and standard logical benchmarks

B. Ablation Results

Table 4: Ablation results under different configurations.

Setting	ProofW.	FOLIO	ProverQA	RepublicQA*	Avg
w/o Square	65.17	72.06	56.60	76.50	67.58
w/o Plan	62.17	69.61	**75.00**	72.00	69.70
w/o Reflect	67.50	76.12	63.40	78.50	71.38
Ours	**71.95**	**79.90**	68.60	**82.50**	**75.74**



C. Semantic Difficulty / Reasoning Hops

Hop↑	FOLIO Acc.↓	FOLIO FK	ProofWriter Acc.↓	ProofWriter FK	ProverQA Acc.↓	ProverQA FK	RepublicQA Acc.↓	RepublicQA FK↑
4	0.5161	6.45	0.6029	1.65	0.5636	8.27	0.4286	12.99
5	0.4474	7.72	0.4634	1.60	0.4231	8.40	0.3125	14.68
6	0.3750	7.35	0.4000	1.38	0.2500	7.40	0.0000	17.70

Finding 2: Semiotic Square is crucial in LogicAgent

Finding 3: Semantic difficulty amplifies logical difficulty

WeChat

Homepage

★ Open to collaboration and communication– feel free to contact me!

ikostar@hust.edu.cn

https://bladey1.github.io/

Wuhan, China