How often does Natural Logic actually meet Machine Learning?

Lasha Abzianidze
ILS, Utrecht University

Disclaimer

Coupling deep learning with natural logic still remains at the heart of NALOMA, a highly promising direction:

Natural logic, a logic with formulas close to natural language forms.

LLMs showing their remarkable capabilities on language input.

The name of NALOMA is not understood literally as the workshop includes the broader topics.

However, the title of this talk should be understood literally.

Surveying works starting from 2023 that uses a sort of natural logic and couple it with machine learning.

Restricted domain: *ACL major venues

Diagram

NatLog
(MacCartney 2009)

NaturalLI
(Angeli&Manning, 2014)

ProofVer
(Krishna et al. 2022)

AdMIRaL
(Aly&Vlachos 2022)

QA-NatVer
(Aly et al. 2023)

TabVer
(Aly&Vlachos 2024)

Zero-NatVer
(Strong et al. 2024)

Natural Logic & Dynamic Rewards for Entailment Tree Gen.
(Shi et al. 2025)

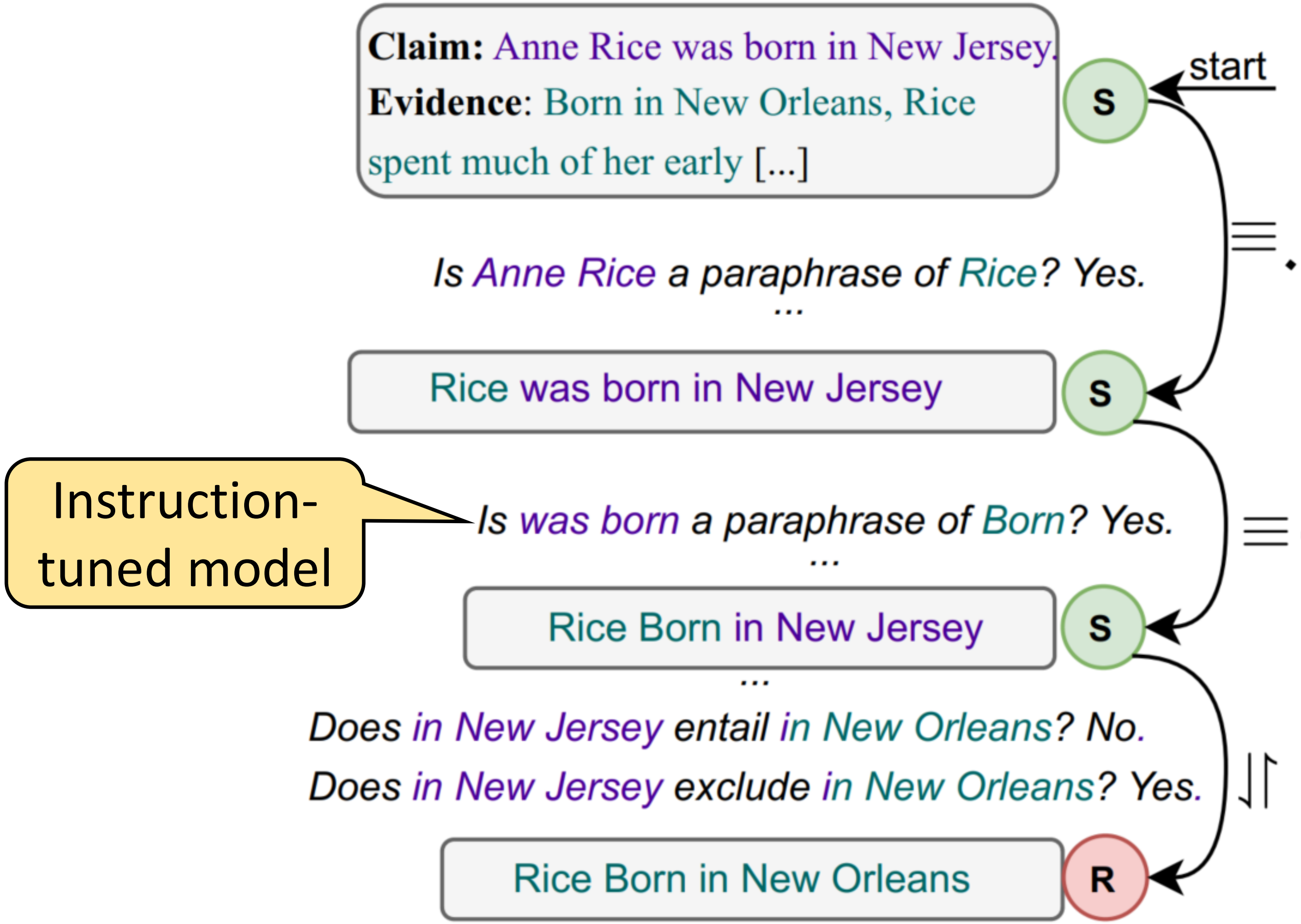
NatLogAttack
(Zheng&Zhu 2023)

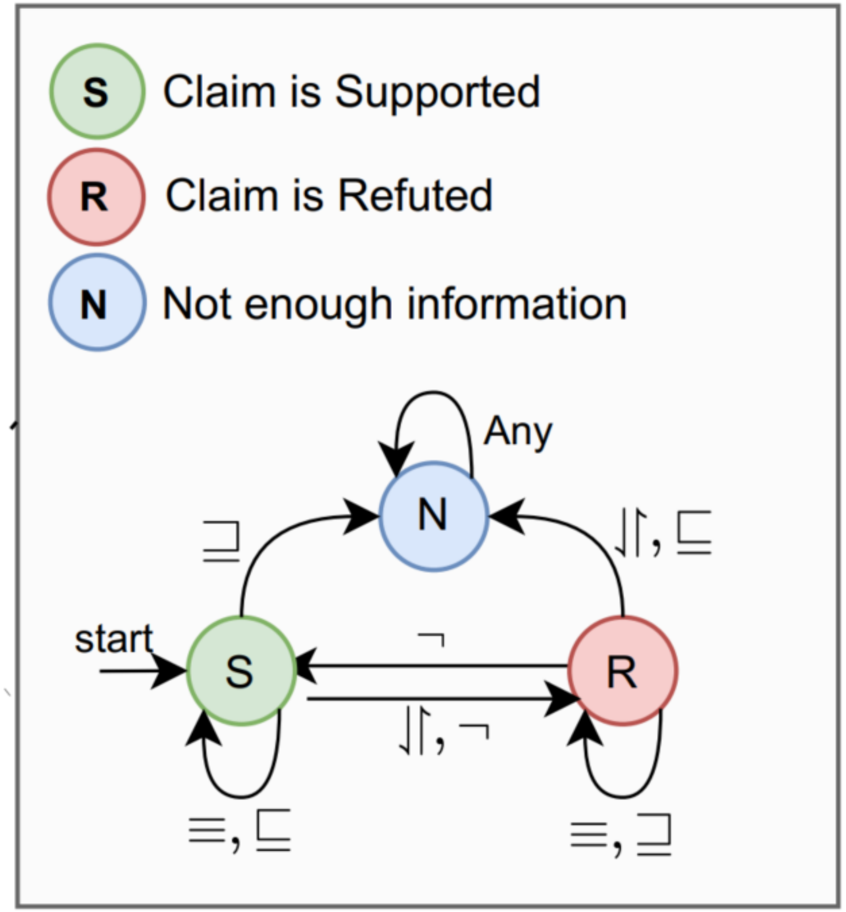
Causal Effect of Monotonicity on NLI models
(Rozanova et al. 2024)

QA-NatVer: Question Answering for Natural Logic-based Fact Verification

Rami Aly, Marek Strong, Andreas Vlachos (University of Cambridge) EMNLP 2023

Angeli&Manning (2014)





Deterministic Finite State Automaton

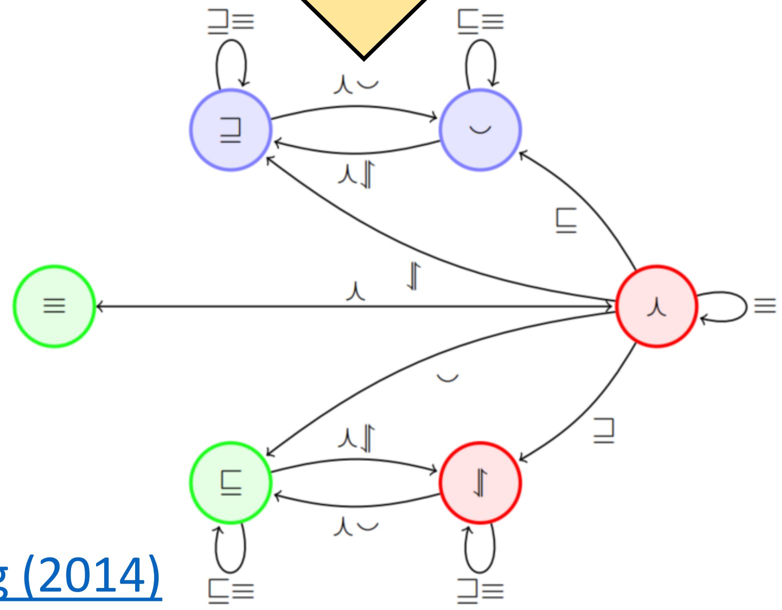
Joining semantic relations

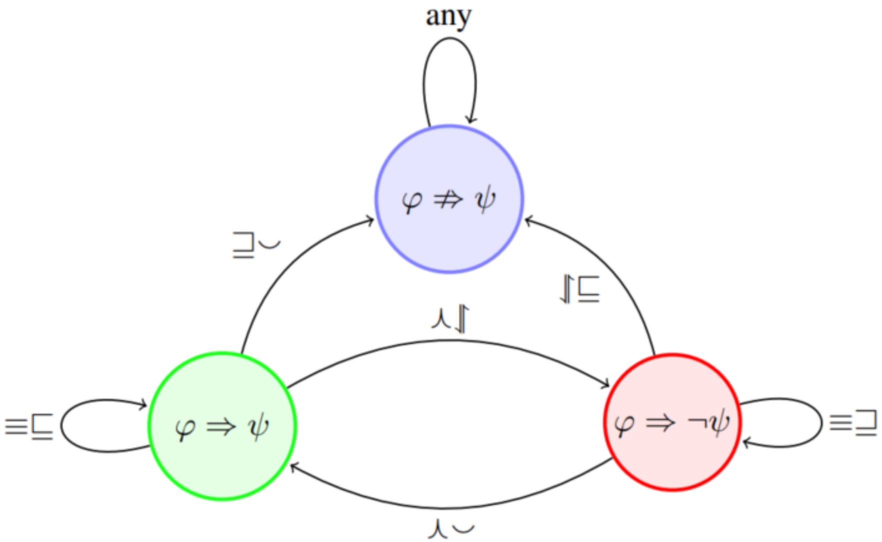
[Icard (2012)](#)

[MacCartney (2009)](#)

$\bowtie$	$\equiv$	$\sqsubset$	$\sqsupset$	$\wedge$	$\mid$	$\smile$	$\#$
$\equiv$	$\equiv$	$\sqsubset$	$\sqsupset$	$\wedge$	$\mid$	$\smile$	$\#$
$\sqsubset$	$\sqsubset$	$\sqsubset$	$\equiv\sqsubset\sqsupset\mid\#$	$\mid$	$\mid$	$\sqsubset\wedge\mid\smile\#$	$\sqsubset\mid\#$
$\sqsupset$	$\sqsupset$	$\equiv\sqsubset\sqsupset\smile\#$	$\sqsupset$	$\smile$	$\sqsupset\wedge\mid\smile\#$	$\smile$	$\sqsupset\smile\#$
$\wedge$	$\wedge$	$\smile$	$\mid$	$\equiv$	$\sqsupset$	$\sqsubset$	$\#$
$\mid$	$\mid$	$\sqsubset\wedge\mid\smile\#$	$\mid$	$\sqsubset$	$\equiv\sqsubset\sqsupset\mid\#$	$\sqsubset$	$\sqsubset\mid\#$
$\smile$	$\smile$	$\smile$	$\sqsupset\wedge\mid\smile\#$	$\sqsupset$	$\sqsupset$	$\equiv\sqsubset\sqsupset\smile\#$	$\sqsupset\smile\#$
$\#$	$\#$	$\sqsubset\smile\#$	$\sqsupset\mid\#$	$\#$	$\sqsupset\mid\#$	$\sqsubset\smile\#$	$\equiv\sqsubset\sqsupset\wedge\mid\smile\#$

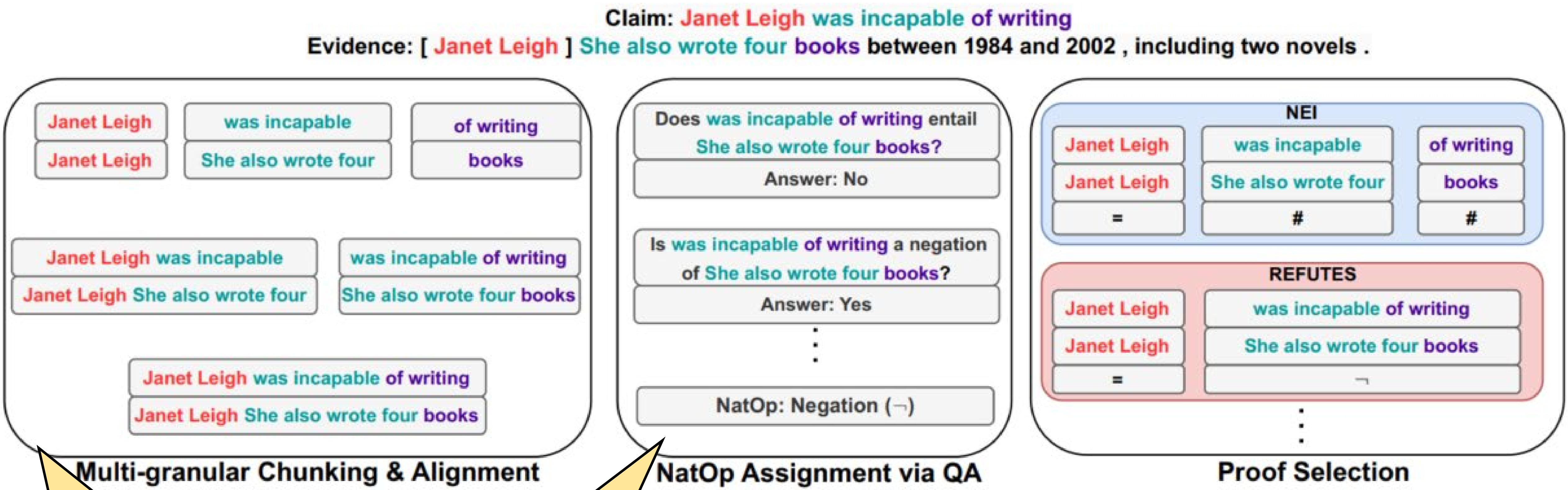
$\bowtie$	$\equiv$	$\sqsubseteq$	$\sqsupseteq$	$\curlywedge$	$\upharpoonleft\downharpoonright$	$\smile$	$\#$
$\equiv$	$\equiv$	$\sqsubseteq$	$\sqsupseteq$	$\curlywedge$	$\upharpoonleft\downharpoonright$	$\smile$	$\#$
$\sqsubseteq$	$\sqsubseteq$	$\sqsubseteq$	$\#$	$\upharpoonleft\downharpoonright$	$\upharpoonleft\downharpoonright$	$\#$	$\#$
$\sqsupseteq$	$\sqsupseteq$	$\#$	$\sqsupseteq$	$\smile$	$\#$	$\smile$	$\#$
$\curlywedge$	$\curlywedge$	$\smile$	$\upharpoonleft\downharpoonright$	$\equiv$	$\sqsupseteq$	$\sqsubseteq$	$\#$
$\upharpoonleft\downharpoonright$	$\upharpoonleft\downharpoonright$	$\#$	$\upharpoonleft\downharpoonright$	$\sqsubseteq$	$\#$	$\sqsubseteq$	$\#$
$\smile$	$\smile$	$\smile$	$\#$	$\sqsupseteq$	$\sqsupseteq$	$\#$	$\#$
$\#$	$\#$	$\#$	$\#$	$\#$	$\#$	$\#$	$\#$





[Angeli&Manning (2014)](#)

QA-NatVer: proof search



Trained word-level aligner

Instruction-tuned model

QA-NatVer: results

Beats the previous work, ProofVer. (Krishna et al., 2022)

Highly efficient and robust, with top performance on the adversarial FEVER.

Suitable for low-resource scenario: outperforms all baselines on the Danish FEVER.

QA-NatVer produces better proofs then ProofVer.

	Accuracy	Macro-Avg. F_1
DeBERTa	39.1 ± 2.5	36.7 ± 2.3
DEBERTa-NLI	59.7 ± 1.9	59.4 ± 1.6
T-Few	59.4 ± 3.2	58.8 ± 3.1
LOREN	35.7 ± 1.8	27.9 ± 1.7
ProoFVer	36.2 ± 1.3	32.5 ± 1.3
QA-NatVer	64.0 ± 0.9	63.1 ± 1.5
+ Flan-T5	$\mathbf{70.3 \pm 2.1}$	$\mathbf{70.0 \pm 1.4}$

On FEVER (32 few shot)

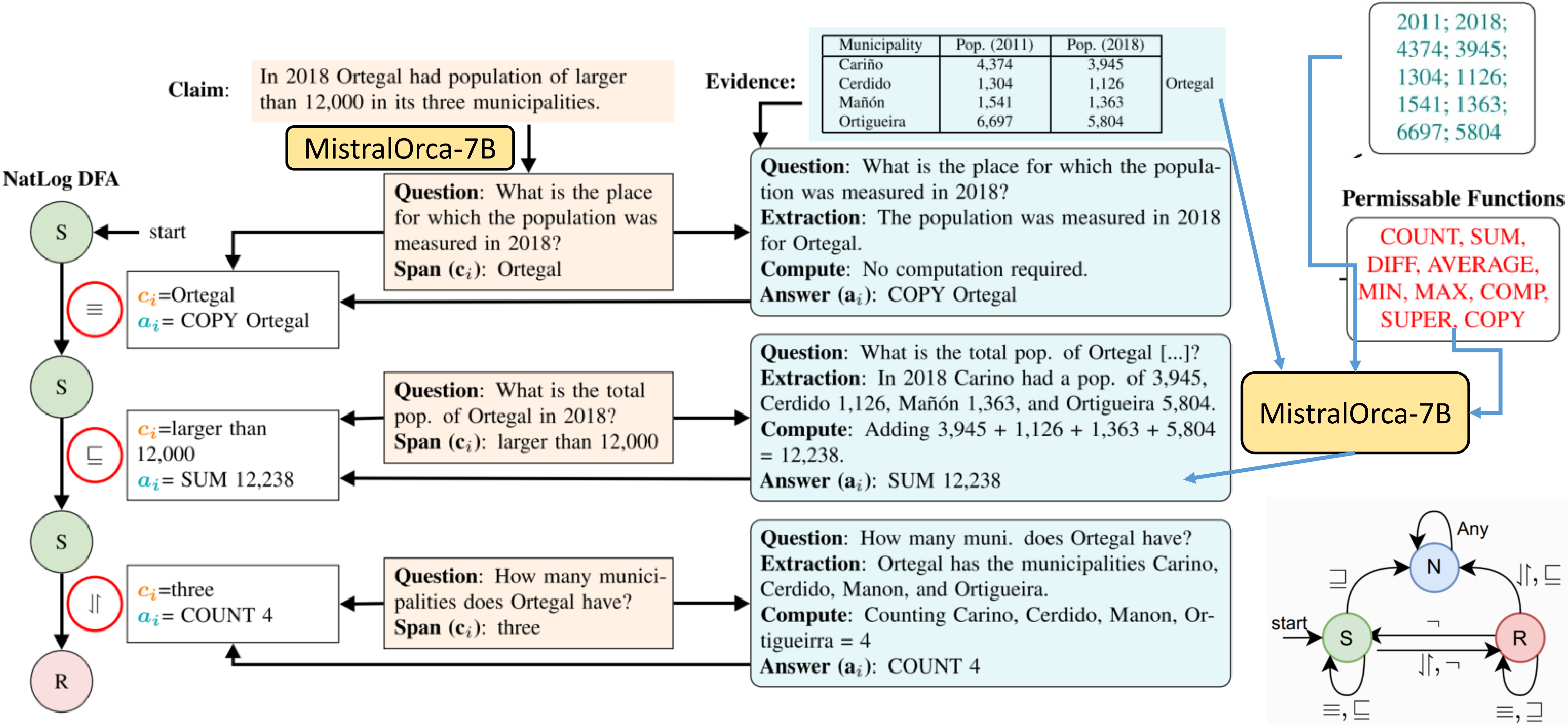
NaturalLI-based

Improvement: **Zero-NatVer** (Strong et al. 2024)

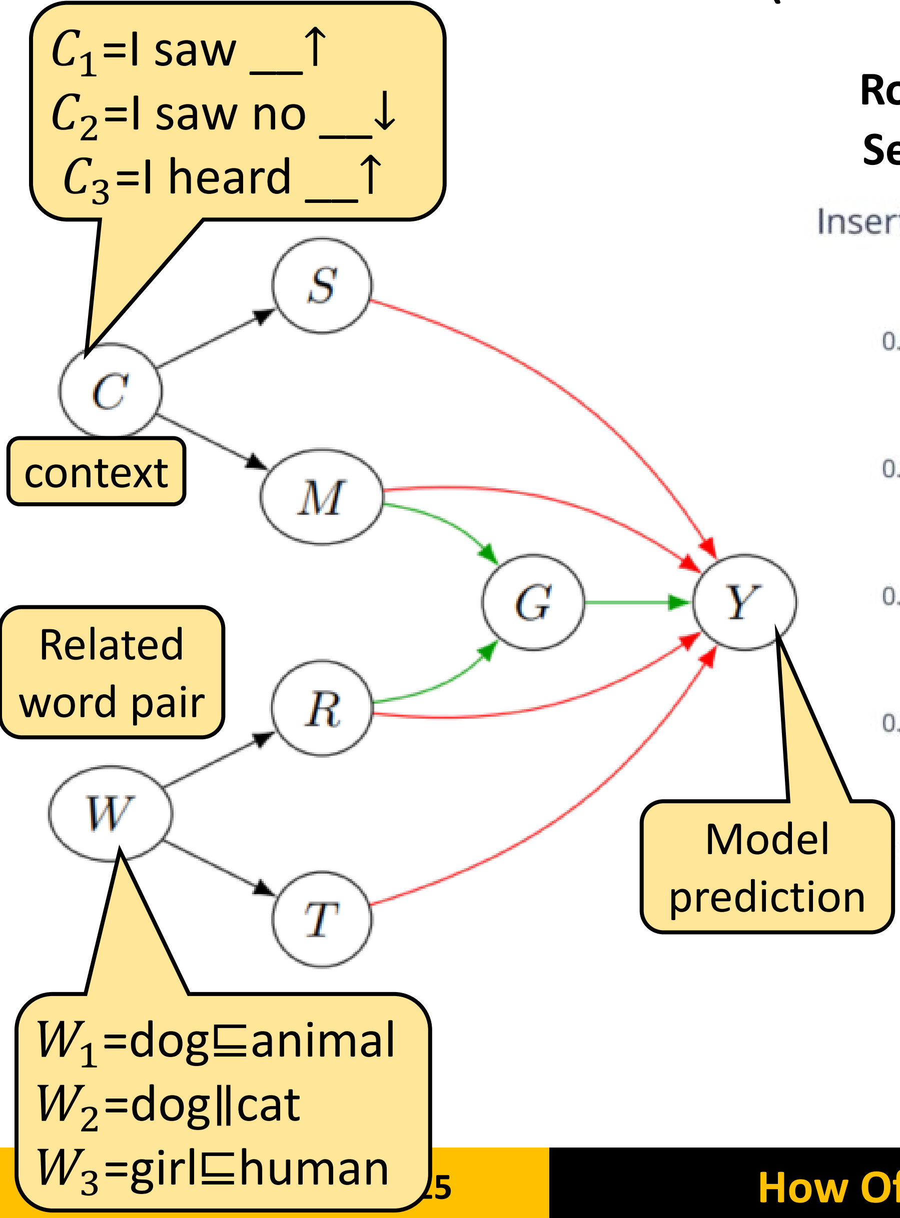
TabVer: Tabular Fact Verification with Natural Logic

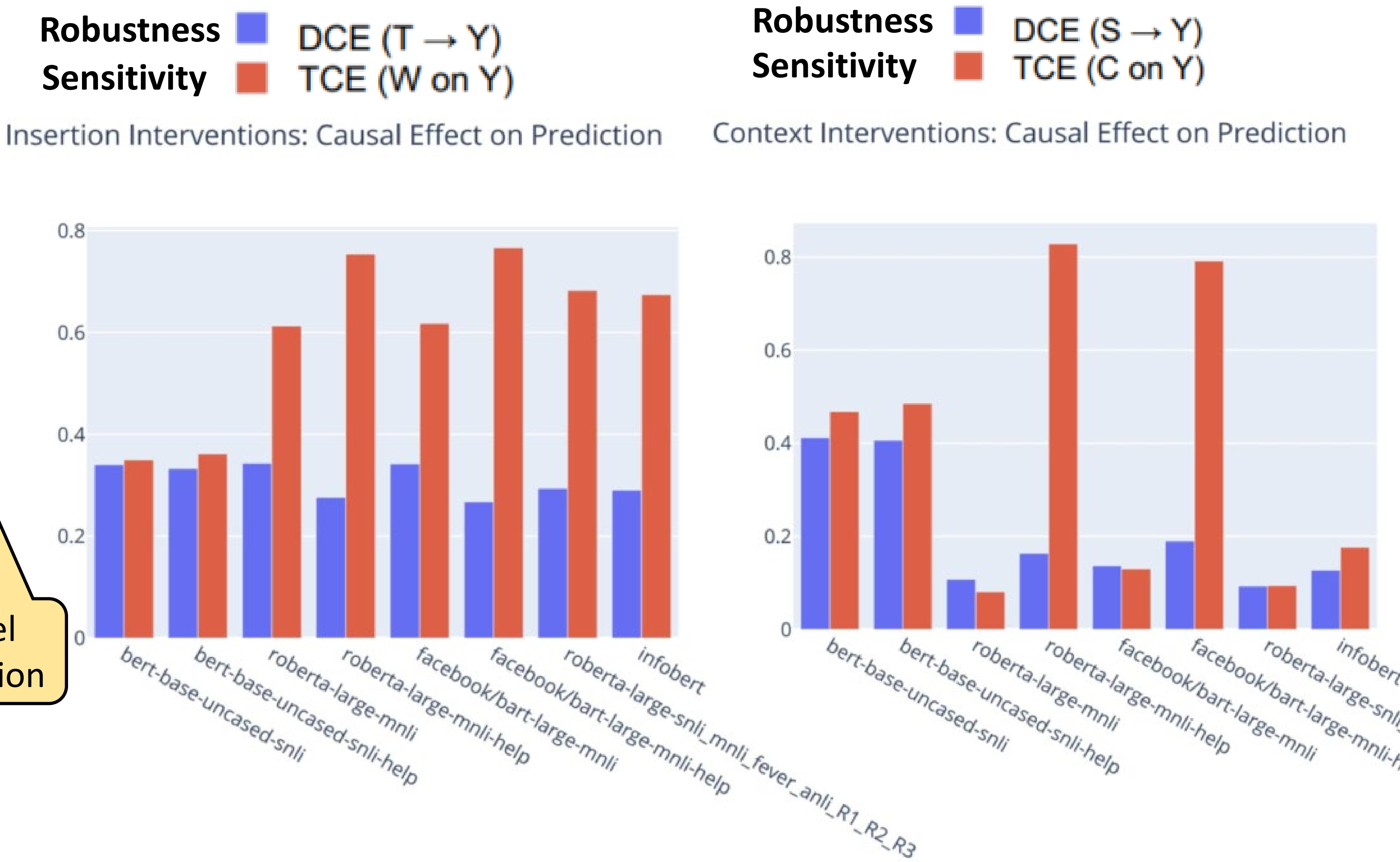
Rami Aly and Andreas Vlachos (University of Cambridge) [TACL 2024]

Permissable Numbers



Estimating the Causal Effects of Natural Logic Features in Transformer-Based NLI Models (Julia Rozanova, Marco Valentino, André Freitas LREC-2024)





Diagram

NaturalLI
(Angeli&Manning, 2014)

NatLog
(MacCartney 2009)

ProofVer
(Krishna et al. 2022)

AdMIRaL
(Aly&Vlachos 2022)

QA-NatVer
(Aly et al. 2023)

TabVer
(Aly&Vlachos 2024)

Zero-NatVer
(Strong et al. 2024)

NatLogAttack
(Zheng&Zhu 2023)

Natural Logic & Dynamic Rewards for Entailment Tree Gen.
(Shi et al. 2025)

Causal Effect of Monotonicity on NLI models
(Rozanova et al. 2024)

How Often does Natural Logic Actually Meet Machine Learning?

Short answer: fairly often…

Long answer: It is important to notice that natural logic is finding its own way in:

- The downstream tasks such as Fact Verification (incl. tabular data) and Question Answering
- The reasonings tasks that require multiple steps and structured explanations.

Hopefully soon:

- Building structured explanations benchmarks with wider-coverage
- (partially) Evaluating faithfulness of GenAI output
- Easier to use Natural Logic-based systems
- …

Closing NALOMA

Reflection

Natural language inference:

In the mood of inference (LLM+FolProver), Ja-NLI for Comparatives,
Visual entailment with synthetic data

ACL Anthology

Natural language understanding tasks:

Legal Reasoning with LLMs
Keynote by Aaron on modeling events/situations

The proceedings will be prepared and injected in the ACL anthology (likely in August)

Evaluation & resources:

Math corpus, MERGE (generalization test for NLI),
Linguistic validity eval for Ja CCGTreebank

Human-AI alignment:

Keynote by Mehrnoosh on next category prediction like humans (GP)
Keynote by Kyle on modeling the logic of GenAI optimization

Next NALOMA

We aim to have next NALOMA in 2026

Co-chairs: Hitomi Yanaka and Lasha Abzianidze

Host venue TBD:

- ESSLLI 2026?
- LREC 2026?
- As *ACL workshop?
- …

Theme of the workshop: generally the same, but maybe with some practical additions such as (un)shared task/challenge.

Program committee

Valeria de Paiva (co-chair), Topos Institute

Lasha Abzianidze (co-chair), Utrecht University

Katrin Erk, University of Texas at Austin

Stergios Chatzikyriakidis, University of Crete

Lawrence S. Moss, Indiana University

Aikaterini-Lida Kalouli, Bundesdruckerei GmbH

Hitomi Yanaka, University of Tokyo and Riken Institute

Hai Hu, Shanghai Jiao Tong University

Thomas Icard, Stanford University

Website: naloma.github.io

@NALOMA_official

@naloma.bsky.social

naloma.pc@gmail.com