

Evaluating Syntactic Structure as a Mechanism for Logical Reasoning in Language Models

Alex Anvi Eponon

Centro de Investigación en Computación del IPN (CIC-IPN), Mexico

Luis Ramos

Centro de Investigación en Computación del IPN (CIC-IPN), Mexico

Moein Shahiki-Tash

Centro de Investigación en Computación del IPN (CIC-IPN), Mexico

Ildar Batyrshin

Centro de Investigación en Computación del IPN (CIC-IPN), Mexico

Grigori Sidorov

Centro de Investigación en Computación del IPN (CIC-IPN), Mexico

Abstract

Recent advances in large language models (LLMs) suggest strong apparent reasoning abilities, yet growing evidence indicates that such performance often arises from surface-level pattern matching rather than genuine compositional understanding. Compositionality, the principle that complex meanings are systematically derived from simpler constituents, has long been viewed as essential for human-like reasoning and generalization. Despite this, modern LLM development has largely favored scaling model size and data over explicit structural representations relying on semantic modeling, raising questions about the necessity and efficiency of such approaches for logical reasoning tasks.

In this study, we examine if language models can use explicit syntactic structure as a strong mechanism for logical reasoning. We present SVOMPT (Subject-Verb-Object-Manner-Place- Time), a canonical syntactic framework that breaks down linguistic inputs into structural elements that can be understood. We, first constructed the dataset under the framework, test six models with parameters ranging from 77M to 1.2B from different generations. With an emphasis on CWQ, HotpotQA, and DROP taken from the QDMR benchmark. The model selection for the question decomposition task spans SmoLLM2, Qwen1.5, Qwen2.5, LFM2.5, Flan-T5-small, and Flan-T5-base and the evaluation experiments were done through Fine-tuning, few-shots, and zero-shot settings.

The idea that reasoning performance scales monotonically with model size is called into question by the preliminary findings. The implementation of SVOMPT allows smaller models to significantly close, and in certain situations, surpass, the performance gap, even though large models perform well in pure zero-shot settings. Notably, when given explicit syntactic decomposition, models with up to 80% fewer parameters perform on par with or better than larger "reasoning-oriented" models.

Fine-tuned SVOMPT models further amplify these gains, demonstrating strong and stable improvements across evaluation metrics. These findings provide first empirical evidences that explicit syntactic structure can function as a computationally efficient alternative to brute-force scaling for reasoning-intensive tasks. This work supports a shift toward structurally grounded, interpretable, and resource-efficient approaches to building reasoning-capable language models.

Keywords

Compositional Reasoning, Question Decomposition, Syntactic Decomposition, Parameter Efficiency, LLMs.