# Selective Dependency Transformer (SDT)

## 1. Overview

The Selective Dependency Transformer (SDT) is a structural optimization of the traditional Transformer architecture. It introduces a syntactic segmentation step that simplifies attention by identifying verbs as central nodes and classifying tokens into semantic blocks—such as subjects, objects, and modifiers. This allows sparse attention to occur only between meaningful, functionally dependent components.

## 2. Core Concept

- Each sentence is decomposed into 'semantic blocks', like Subject Block, Object Block, etc.

- Verbs act as hubs—central tokens that define sentence structure and semantic routing.

- Token blocks (phrases or clauses) are encoded separately and interact sparsely via verb-directed attention.

## 3. Semantic Block Structure

| Block Type | Description |

|-------------------|------------------------------------------|

| Verb Node | Core of semantic interaction |

| Subject Block | Initiator of the action (noun/clause) |

| Object Block | Receiver of the action (noun/clause) |

| Modifier Block | Adverbs, adjectives, prepositional info |

| Clause Chain | Embedded clauses acting as arguments |

## 4. Computational Flow

1. Perform syntactic parsing (identify verbs, chunk subjects/objects)

2. Encode each semantic block independently

3. Apply sparse cross-block attention, mainly verb-centered

4. Aggregate semantic representation for downstream tasks

## 5. Advantages

- Reduces attention complexity by 40–60%

- Maintains or improves accuracy through explicit structure

- Compatible with existing Transformers as a preprocessing layer

- Expandable to structured text, code, or multilingual syntax

## 6. Visualization Plan (for PDF v2)

A tree-style diagram showing a sentence like:

'The cat quickly chased the mouse under the sofa.'

- Verb Node: chased

- Subject Block: The cat

- Modifier Block: quickly

- Object Block: the mouse

- Prepositional Block: under the sofa

## 7. Conclusion

SDT introduces grammar-aware, computation-efficient processing to NLP models. It reduces wasteful attention and aligns deep learning models with human language structure. This architecture offers a new path toward smarter, faster language understanding systems.