

Input Processing

Cross-Attention Fusion

Training & Inference

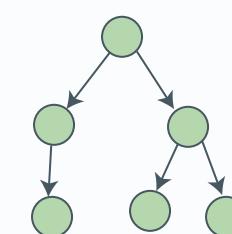
Semantic Encoder

Word Subwords → Character-offset Alignment

The ambience is very pleasant but the price is high

Syntactic Encoder

Syntactic Encoder (Graph Transformer)



Tree Distance Bias
Dependency Type Bias

Aligned Semantic

H'_{sem}

$L \times 768$

$Attn_{sem \leftarrow syn}$

$$Q = H'_{sem} \\ K/V = H_{syn}$$

\tilde{H}'_{sem}

$L \times 768$

MeanPool
MaxPool

Concat

Linear projection

h^{fused}

768

H_{syn}

$L \times 768$

$Attn_{syn \leftarrow sem}$

$$Q = H_{syn} \\ K/V = H'_{sem}$$

\tilde{H}_{syn}

$L \times 768$

MeanPool
MaxPool

Syntactic

Training Branch

Contrastive Learning

| | | High | Low |
|------|----------------------------|-----------------------------|---------------------|
| | | Syntactic similarity | Semantic Similarity |
| High | Semantically Hard Negative | Positive | High |
| Low | Random Negative | Syntactically Hard Negative | Low |

InfoNCE Loss

Inference Branch

Example Retrieval

Cosine Similarity Computation

Top-K Selection

Input: The atmosphere is very inviting but the cost is high.
Output: [{atmosphere, very inviting, POS}, {cost, high, NEG}]
...

LLM Prompt

Structured Predictions