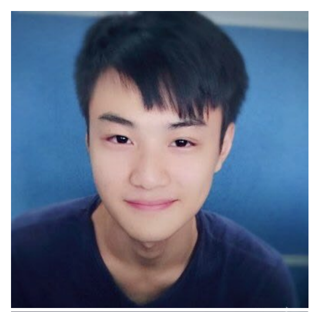
Core Semantic First: A Top-down Approach for AMR Parsing



Deng Cai and Wai Lam

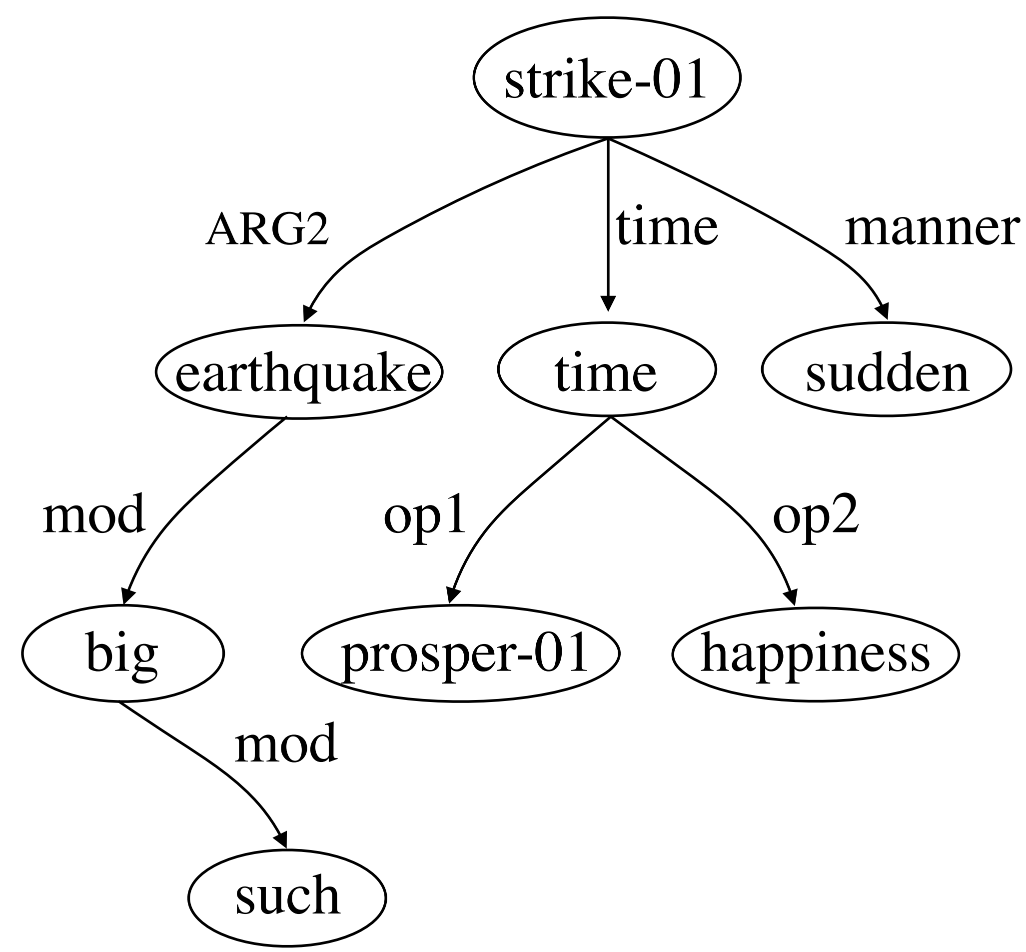
The Chinese University of Hong Kong

EMNLP-IJCNLP2019

Background

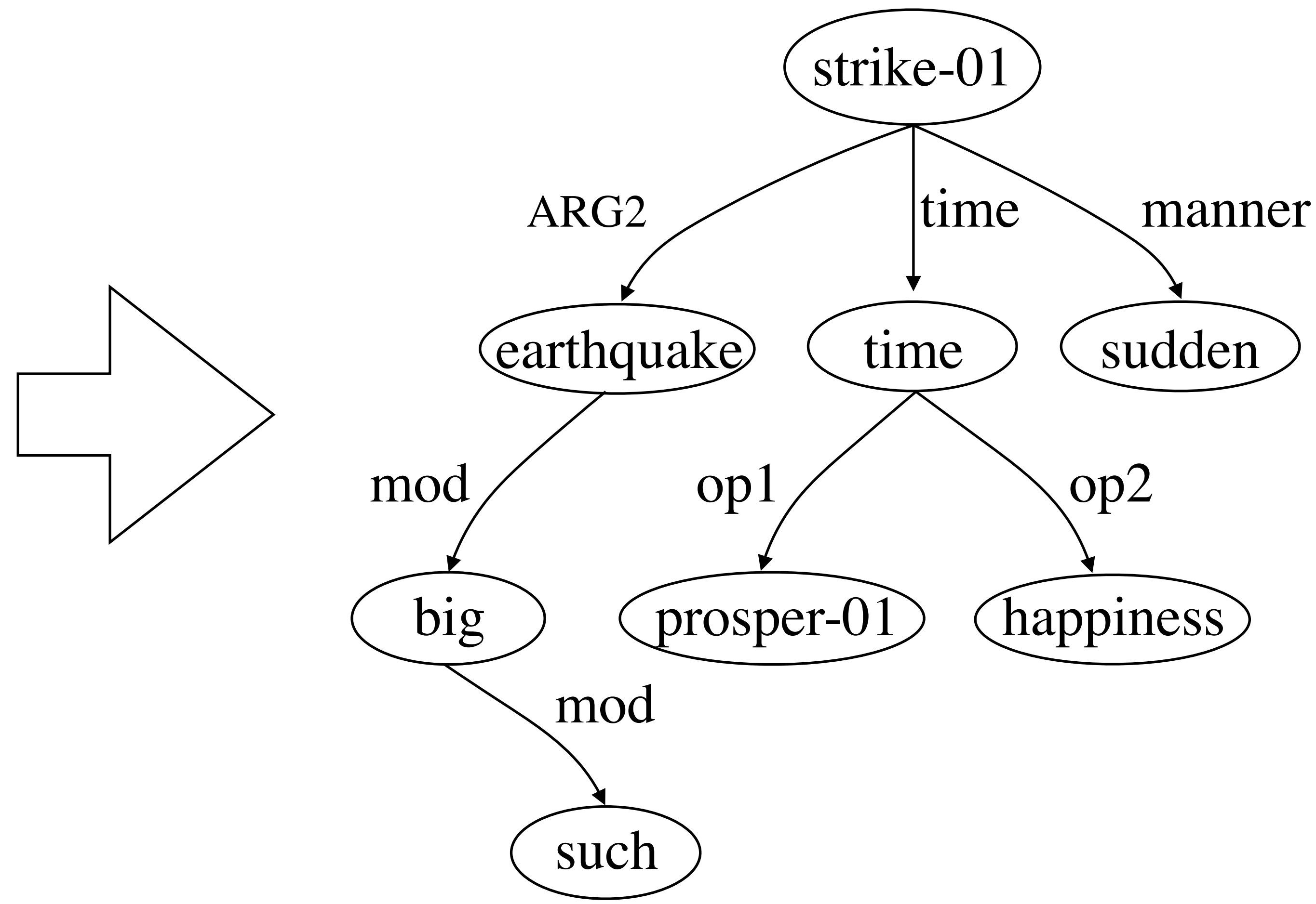
- Abstract Meaning Representation (AMR)
  - rooted, labeled, directed graph
  - nodes represent concepts
  - edges represent relations

During a time of prosperity and happiness, such a big earthquake suddenly struck.



Challenges

During a time of prosperity and happiness, such a big earthquake suddenly struck.



- No explicit alignment of graph nodes and sentence tokens
- Frequent reentrancies and non-projective arcs
- Large and sparse concept vocabulary

Existing Work

- **Graph**-based parsers (Flanigan et al., 2014; Lyu and Titov, 2018, Zhang et al., 2019)
    - pipeline design for concept identification and relation prediction
- **Transition**-based parsers (Wang et al., 2016; Damonte et al., 2017; Ballesteros and Al-Onaizan, 2017; Guo and Lu, 2018; Liu et al., 2018; Wang and Xue, 2017)
    - process a sentence from left-to-right and constructs the graph incrementally
- **Seq2Seq**-based parsers (Barzdins and Gosko, 2016; Konstas et al., 2017; van Noord and Bos, 2017)
    - output a linearization (depth-first traversal) of the AMR graph.

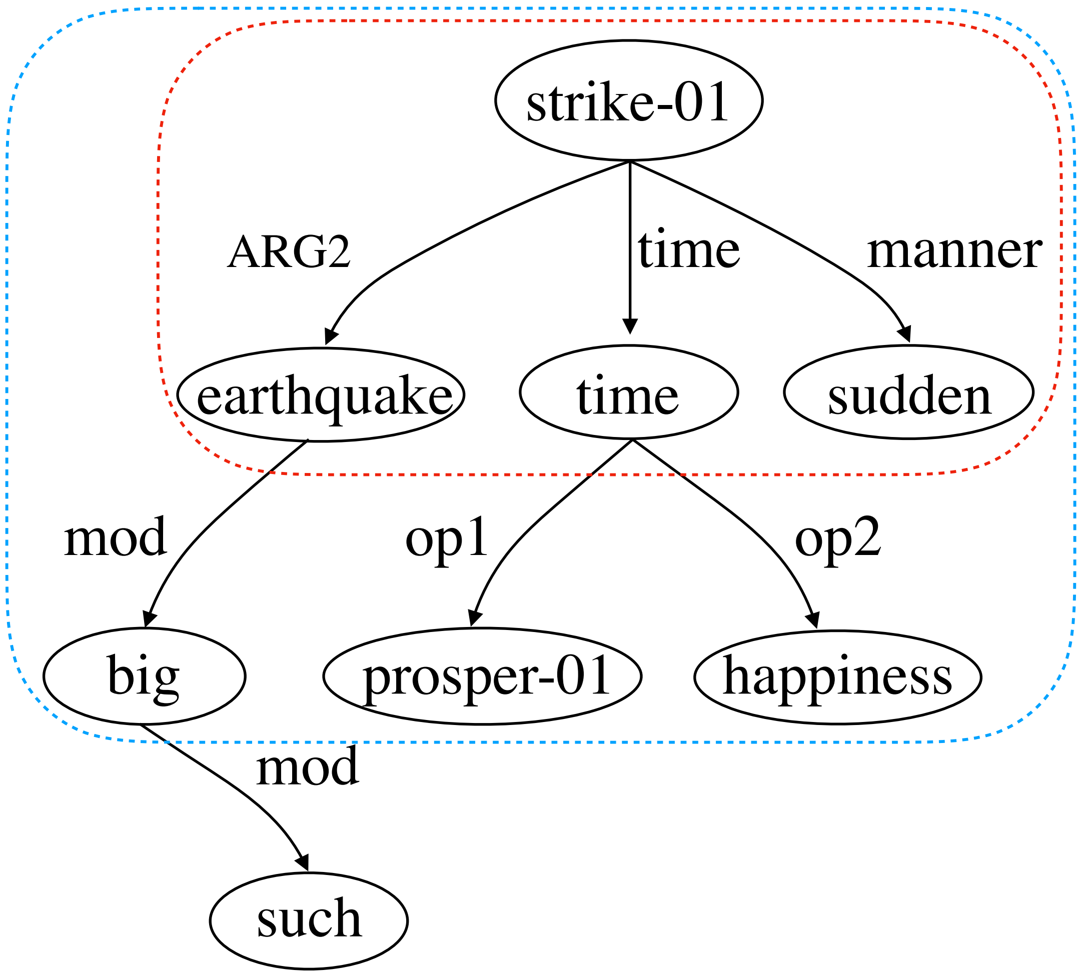
Motivation

- **Graph**-based parsers
  - misses the the interactions between individual decisions
- **Transition**-based & **Seq2Seq**-based parsers
  - suffers from error propagation, where later decisions can easily go awry.

- **Our framework**
  - has a global view and
  - a priority for capturing the main ideas first

Motivation

- **Core Semantic First**



An earthquake suddenly struck at a particular time.

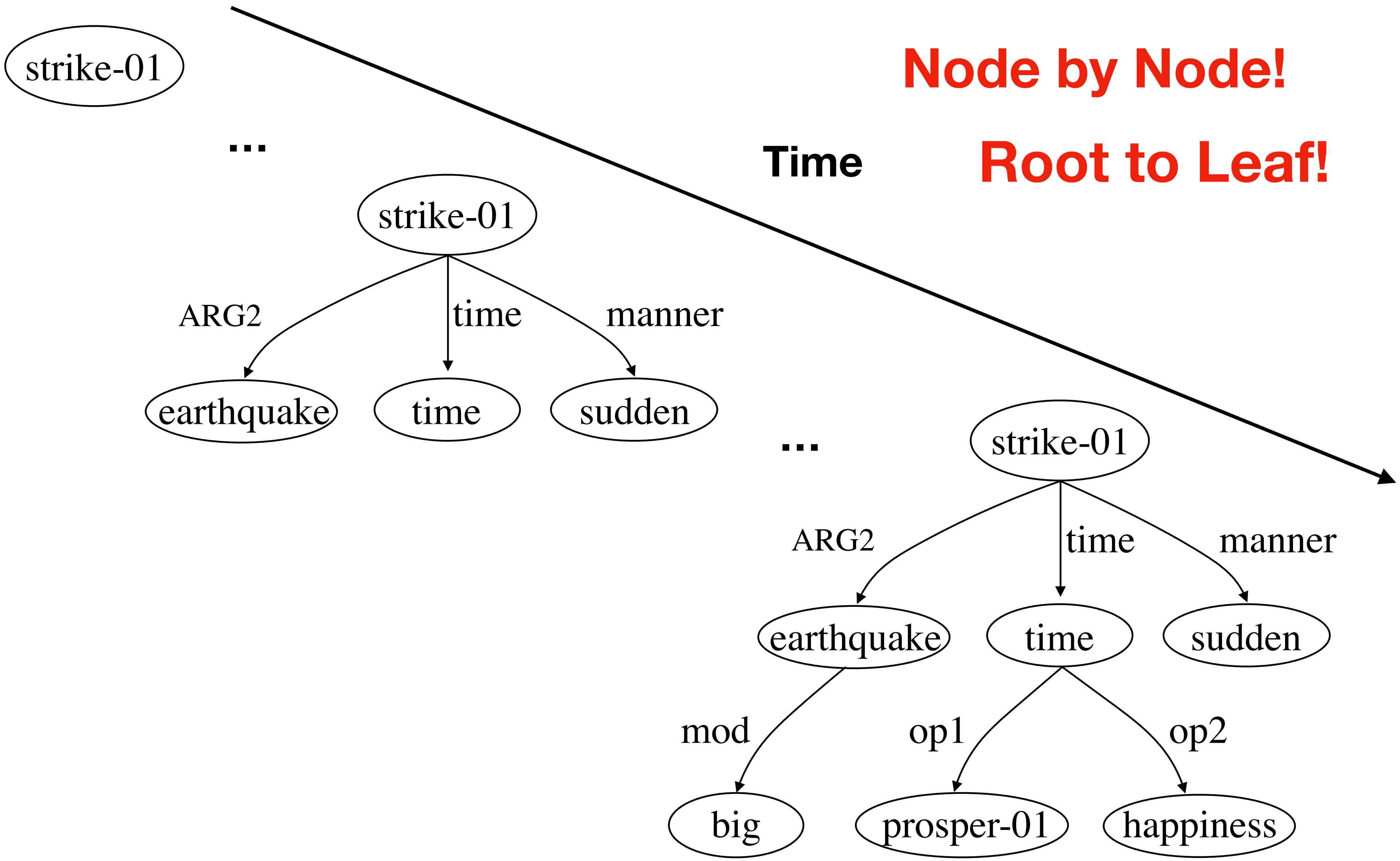
↓

During a time of prosperity and happiness, such a big earthquake suddenly struck.

↓

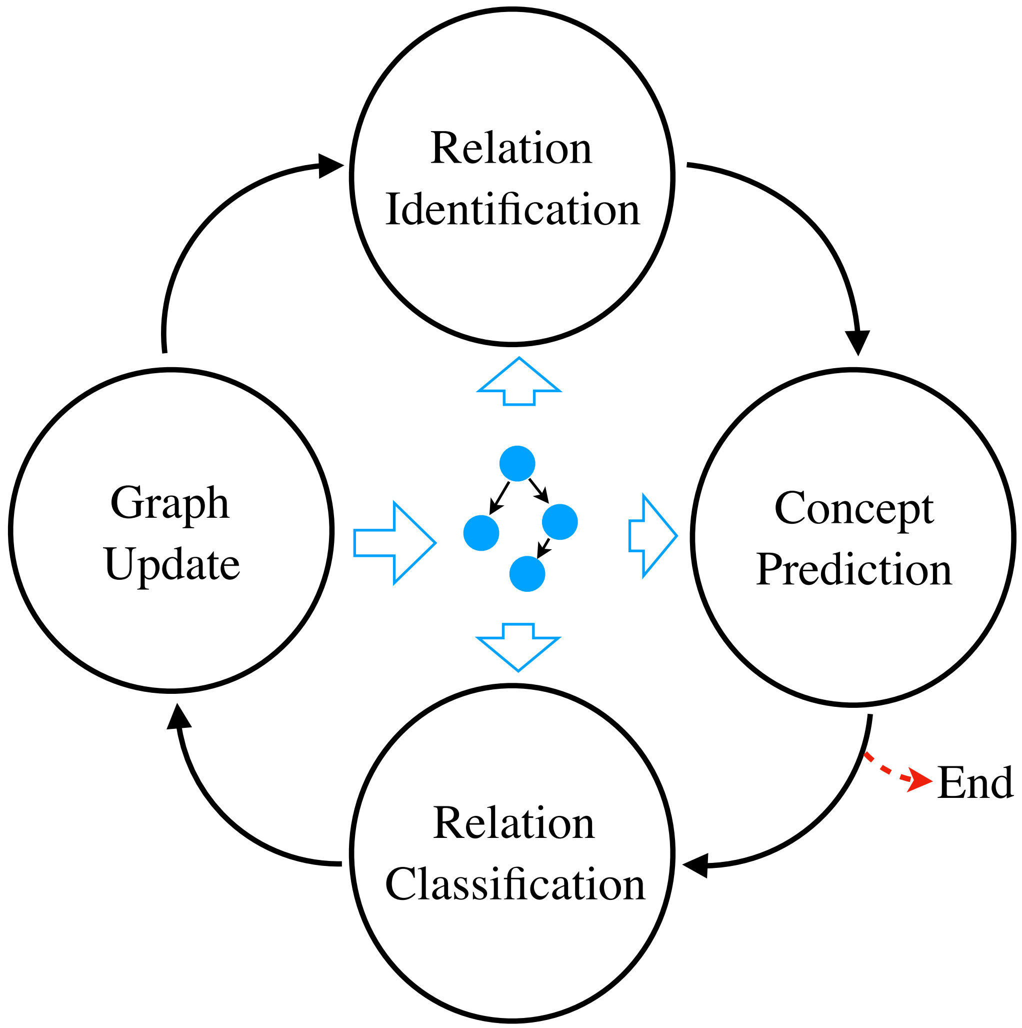
During a time of prosperity and happiness, such a big earthquake suddenly struck.

Graph Spanning



Overview

- Graph spanning-based Parsing
    - starts from the root
    - spans the nodes by the distance to the root
    - at each step, a new node and its connections to existing nodes will be jointly predicted.



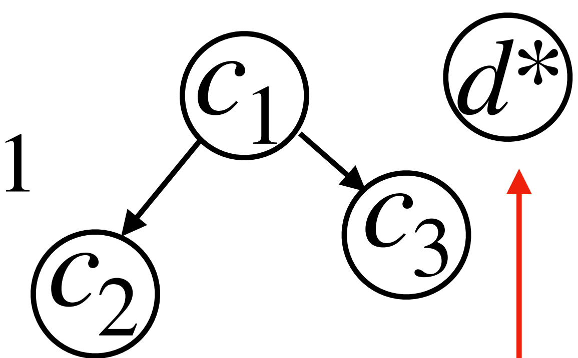
Comparisons

- versus Graph-based Methods:
    - Captures more complicated intra-graph interactions
- versus Transition-based Methods:
    - Removes the left-to-right restriction
    - Avoids sophisticated oracle design for handling the complexity of AMR graphs
- versus Seq2seq-based Methods:
    - Makes direct use of the graph structure information

At time step t

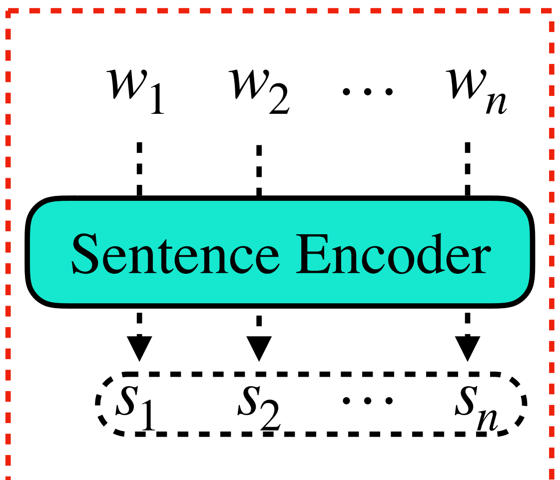
$w_1 \quad w_2 \quad \cdots \quad w_n$

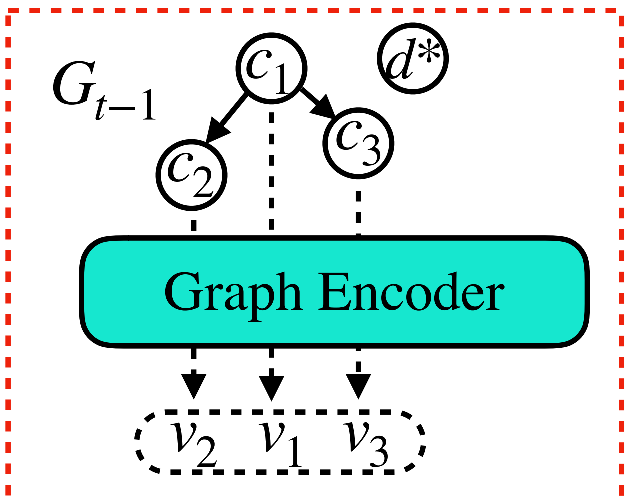
G_{t-1}



A dummy node we used to initialize the graph.

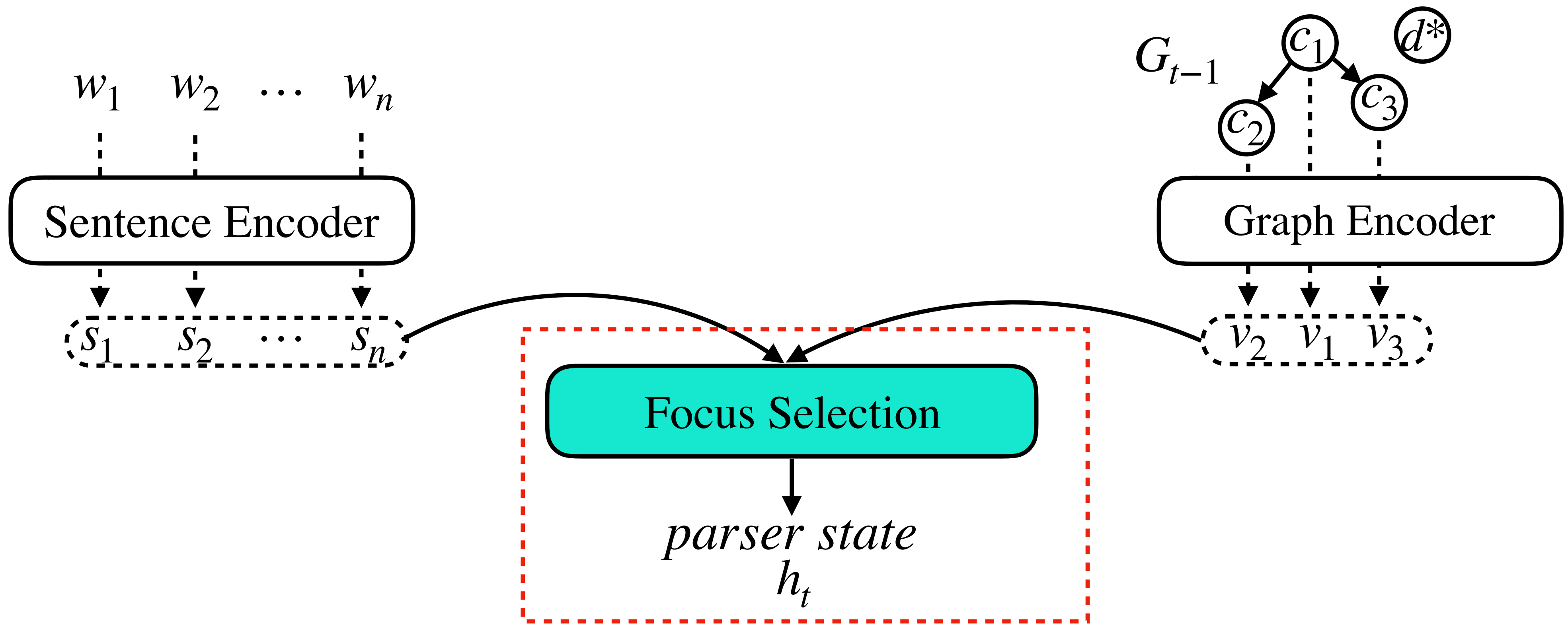
Sentence & Graph Encoders





- Transformer Encoders
- Graph is treated as a sequence of nodes for simplicity.

Focus Selection

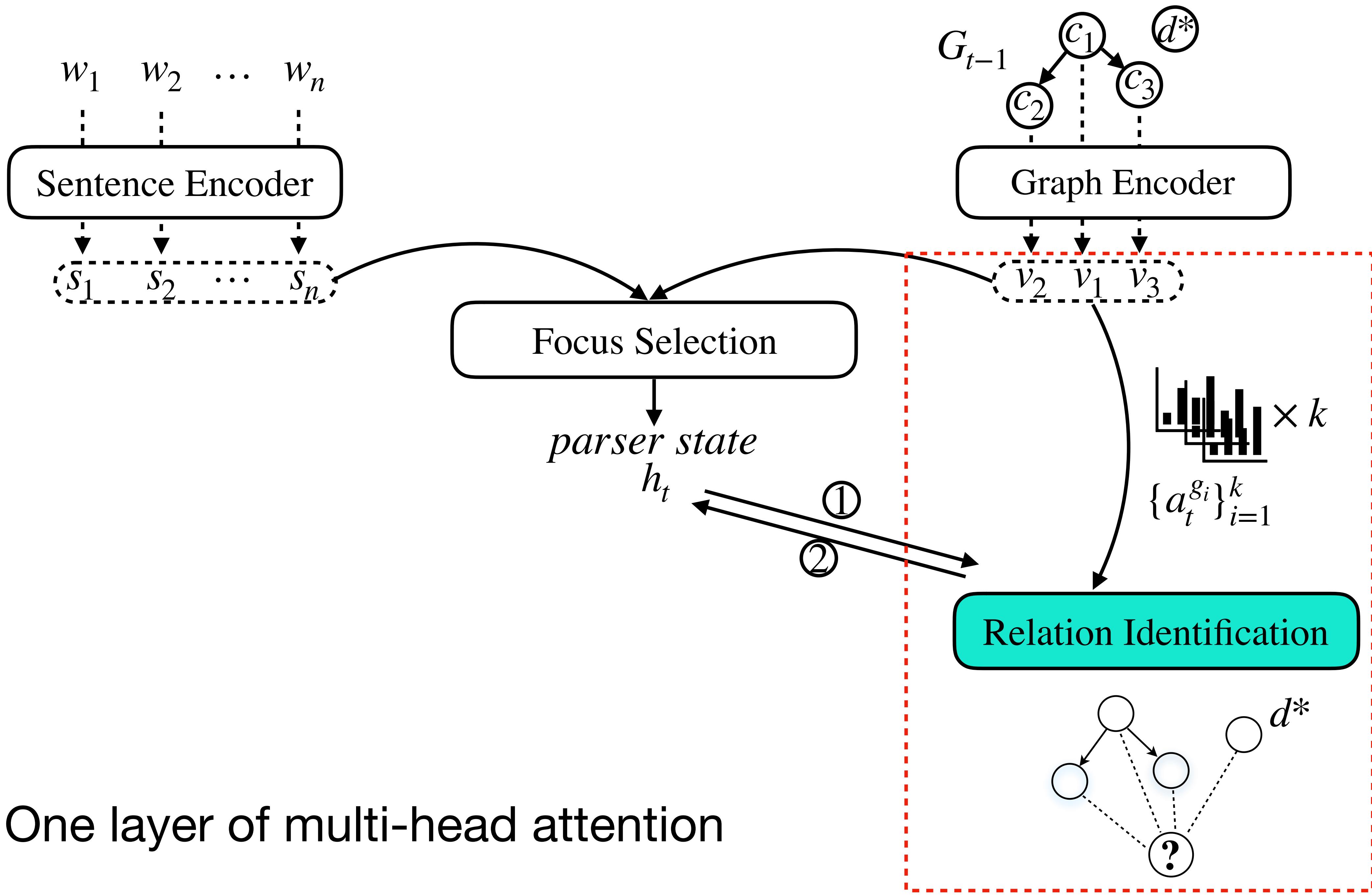


- Multiple layers of multi-head attention

$$x_{t,1}^{l+1} = \mathrm{LN}(h_t^l + T_1^{l+1}(h_t^l, s_{1:n}))$$
$$x_{t,2}^{l+1} = \mathrm{LN}(x_{t,1}^{l+1} + T_2^{l+1}(x_{t,1}^{l+1}, v_{0:t-1}))$$
$$h_t^{l+1} = \max(x_{t,2}^{l+1} W_1^{l+1} + b_1^{l+1}) W_2^{l+1} + b_2^{l+1}$$

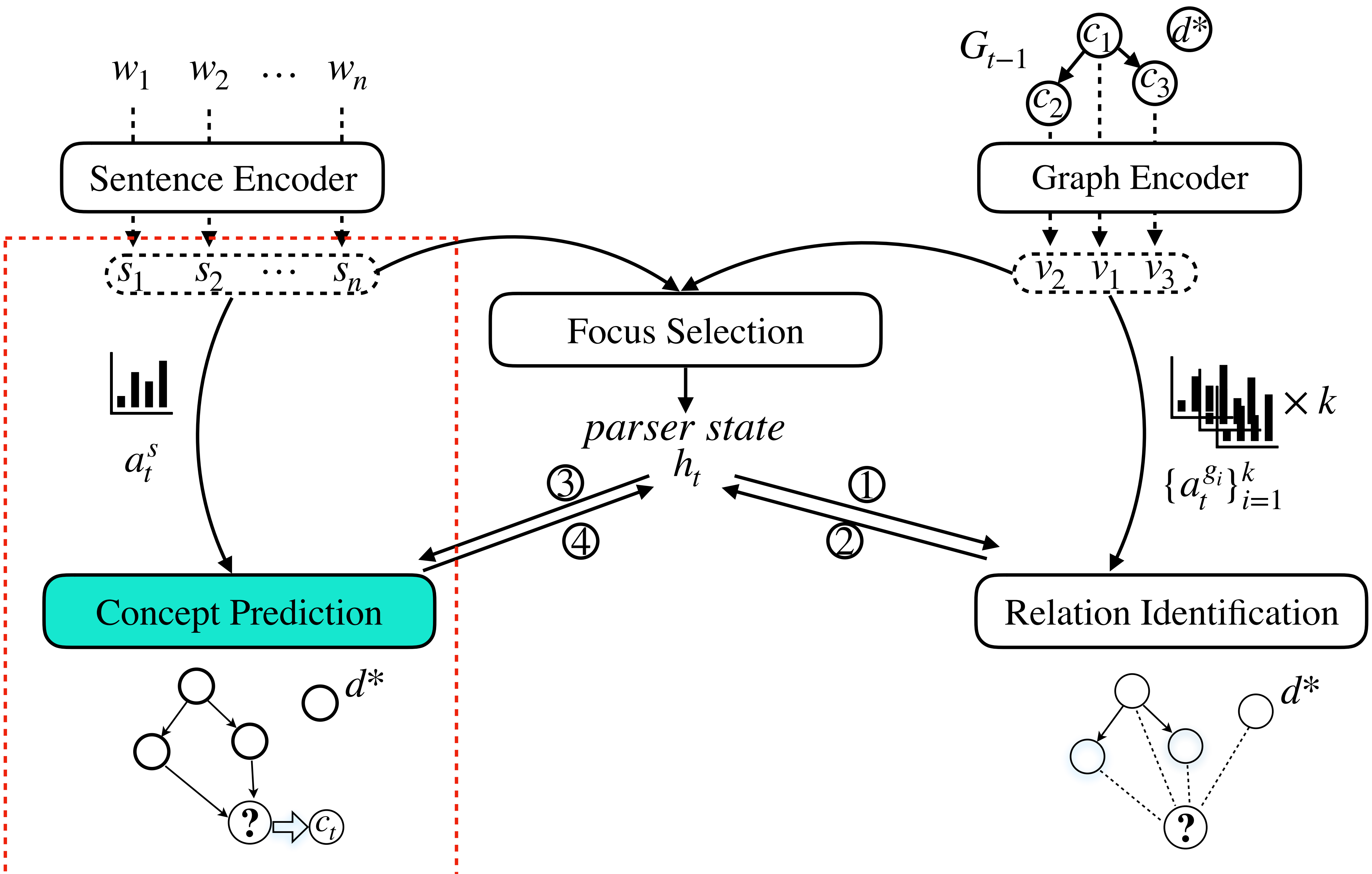
- Collect the most relevant information for the next expansion

Relation Identification

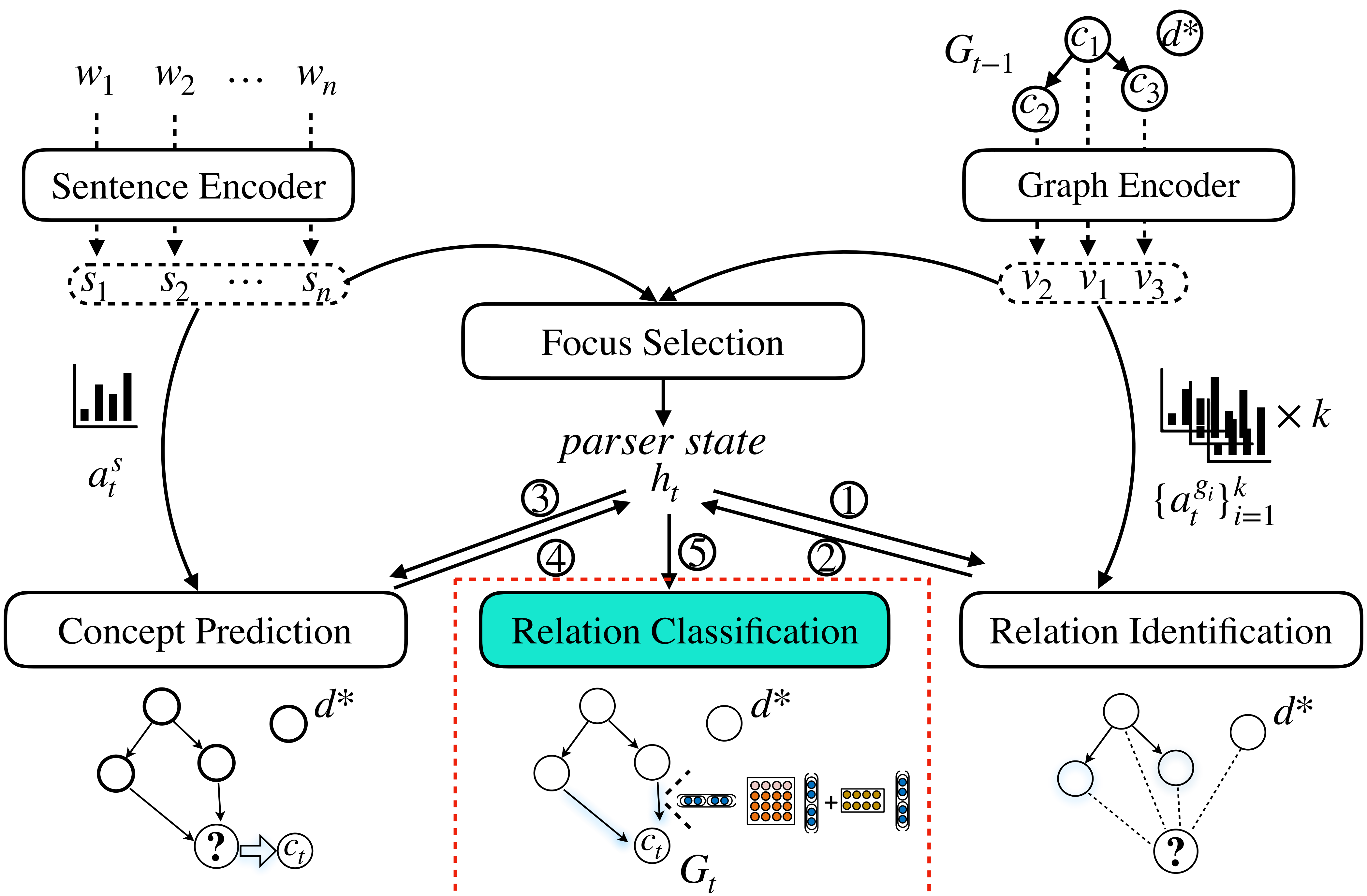


- One layer of multi-head attention
- The maximum over different heads as the final arc probabilities

Concept Prediction

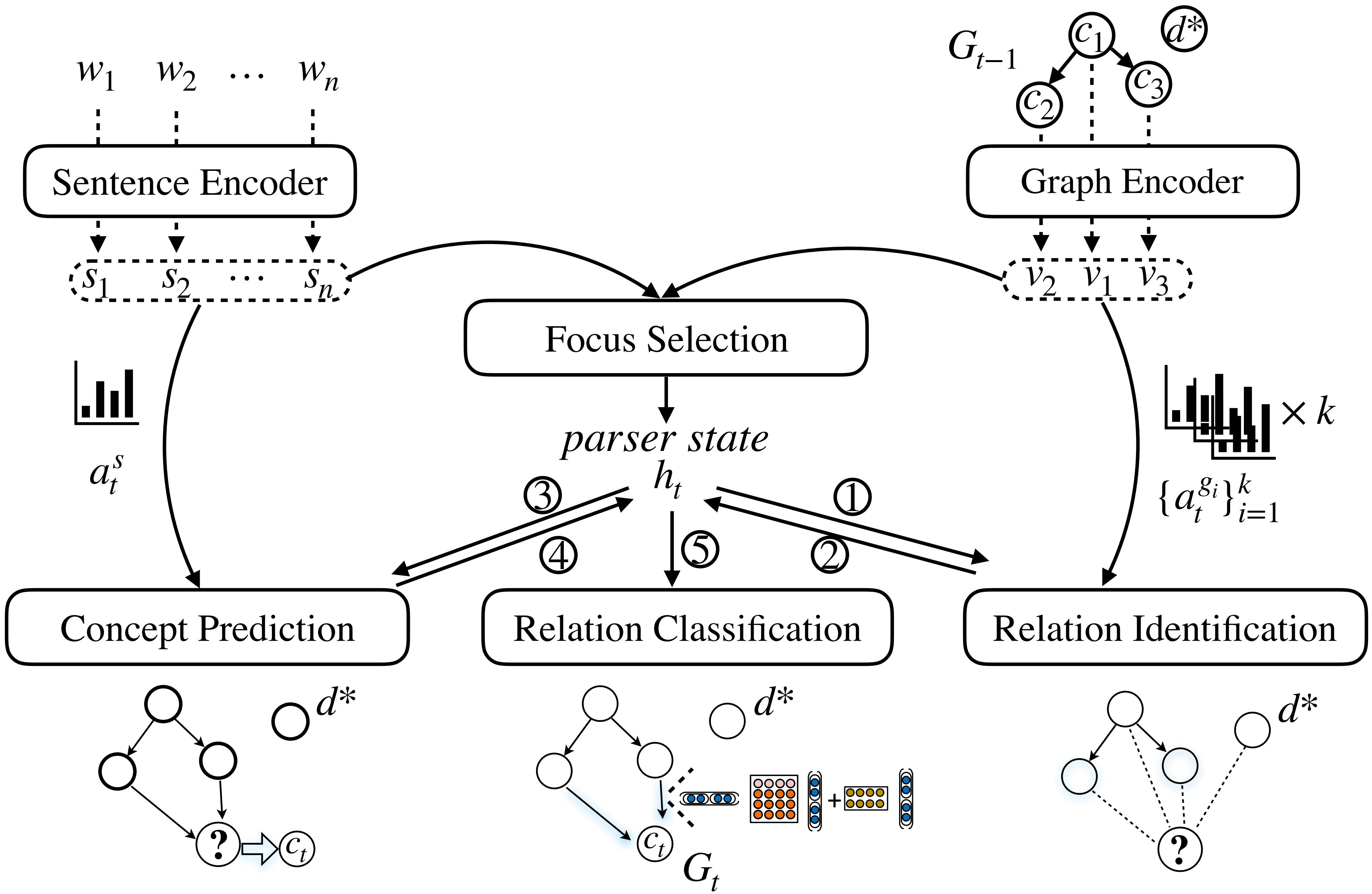


- One layer of single-head attention
- A soft alignment to sentence tokens (also used for copy mechanism)

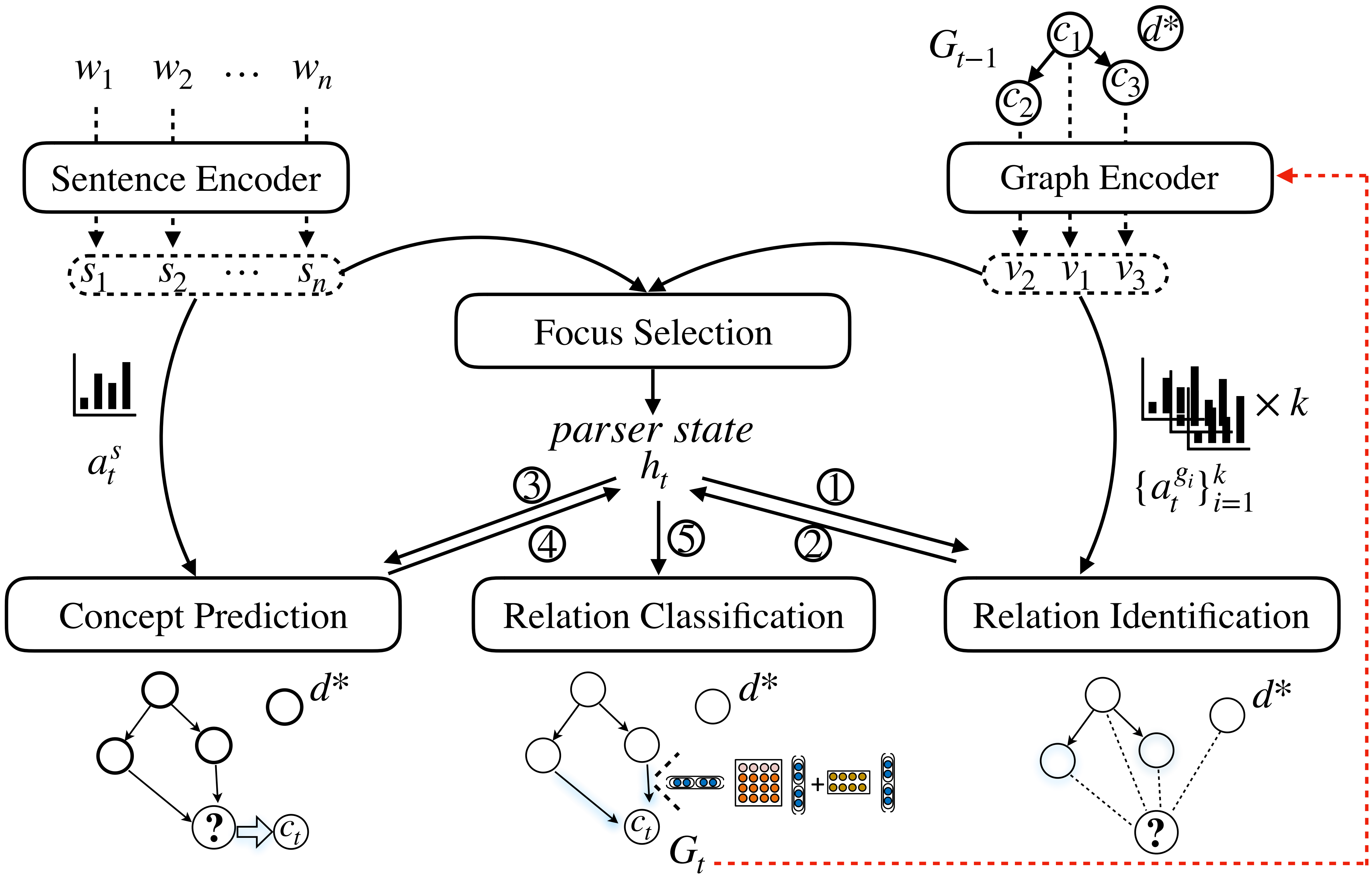


- Biaffine Classifier (Dozat and Manning, 2016)

Architecture



Once Again



w_1 w_2 ... w_n
Sentence Encoder
s_1 s_2 ... s_n
G_{t-1}
c_1 c_2 c_3 $d*$
Graph Encoder
v_2 v_1 v_3
Focus Selection
a_t^s
$\times k$
$\{a_t^{g_i}\}_{i=1}^k$
parser state
h_t
① ② ③ ④ ⑤
Concept Prediction
Relation Classification
Relation Identification
$d*$
?
c_t
$d*$
c_t
G_t
$d*$
?

Training and Inference

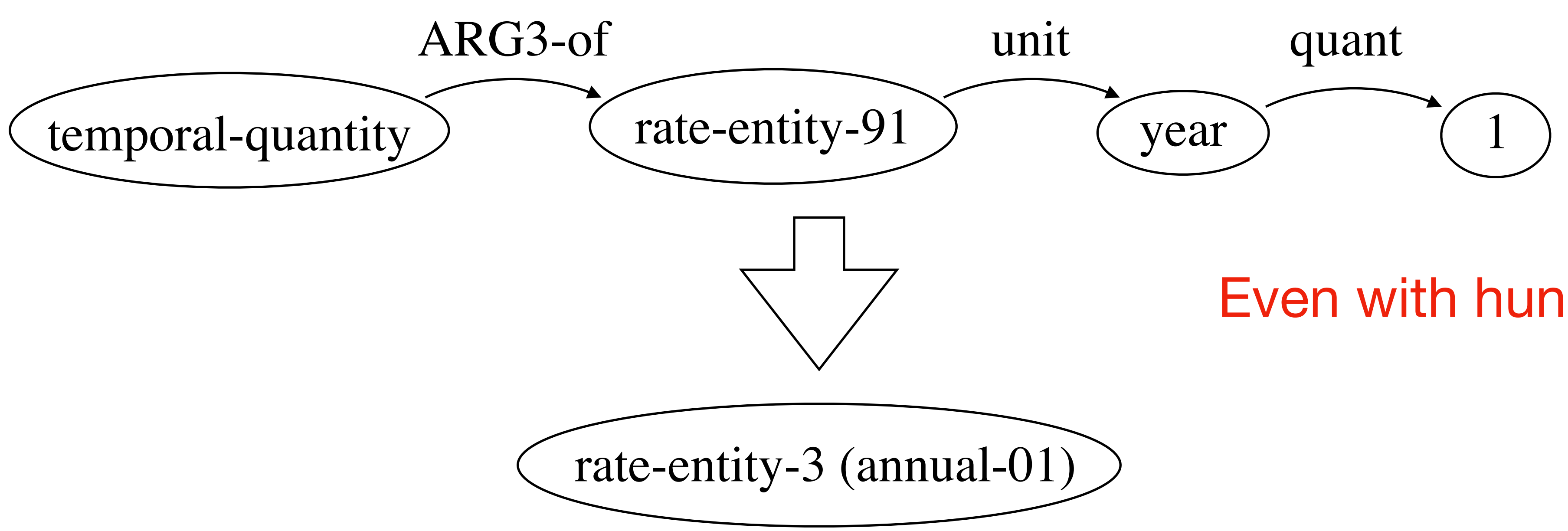
- Autoregressive model
- Distribution Factored according to a top-down graph structure
- Clear separation of node, arc and relation label probabilities
- Beam search (top K graphs)

$$\log P(G|\mathbf{w}) = \sum_{t=1}^{m} \Big(\log P(c_t|G_{t-1}, \mathbf{w}) + \sum_{i \in pred(t)} \log P(arc_{it}|G_{t-1}, \mathbf{w}) + \sum_{i \in pred(t)} \log P(rel_{arc_{it}}|G_{t-1}, \mathbf{w}) \Big)$$

Setup

- The latest AMR sembank (LDC2017T10)
- 36521, 1368, and 1371 sentences in the training, development, and testing sets respectively

Graph Re-categorization



Even with hundreds more!

- Non-trivial. It requires exhaustive screening and expert-level manual efforts.

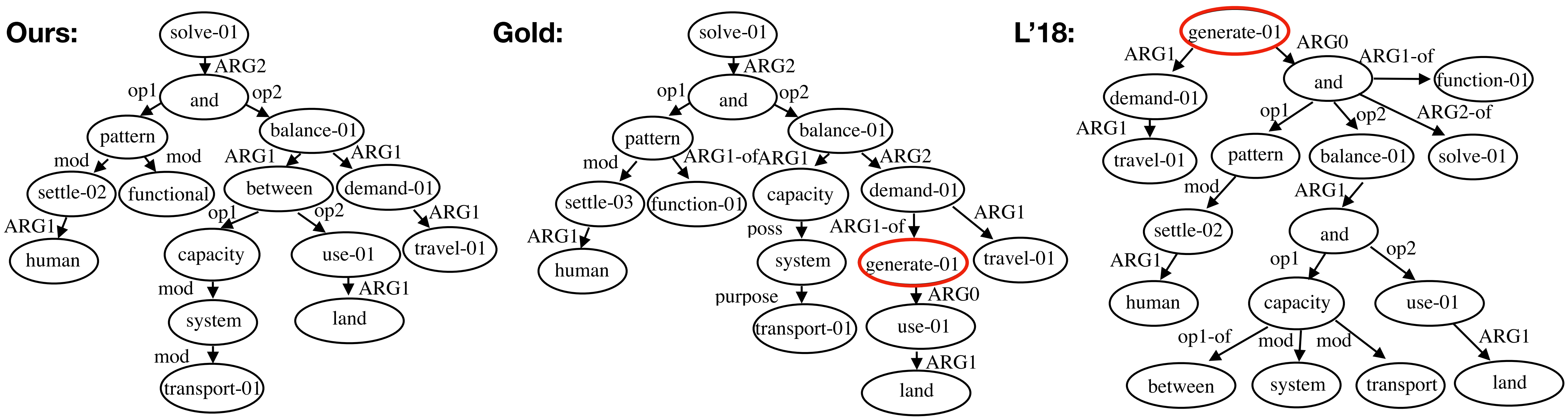
- The precise set of re-categorization rules differs among different models.

Evaluation Metrics

- Smatch (Cai and Knight, 2013) : seeks the maximum overlap after transforming graph into triples.
- Smatch-weighted: assigns more weights to triples stay closer to the root.
- Smatch-core: only compares the subgraphs close to the root.
- Complete-match (CM): completely correct rate
- Root-accuracy (RA): root accuracy

Case Study

Input: *The solution is functional human patterns and a balance between transport system capacity and land use generated travel demand.*



Lyu and Titov, 2018

- Smatch-weighted: 74% vs. 61% Smatch-ordinary: 68% vs. 66%
- The ordinary Smatch is not a proper metric for evaluating the quality of capturing core semantics.

Compared Methods

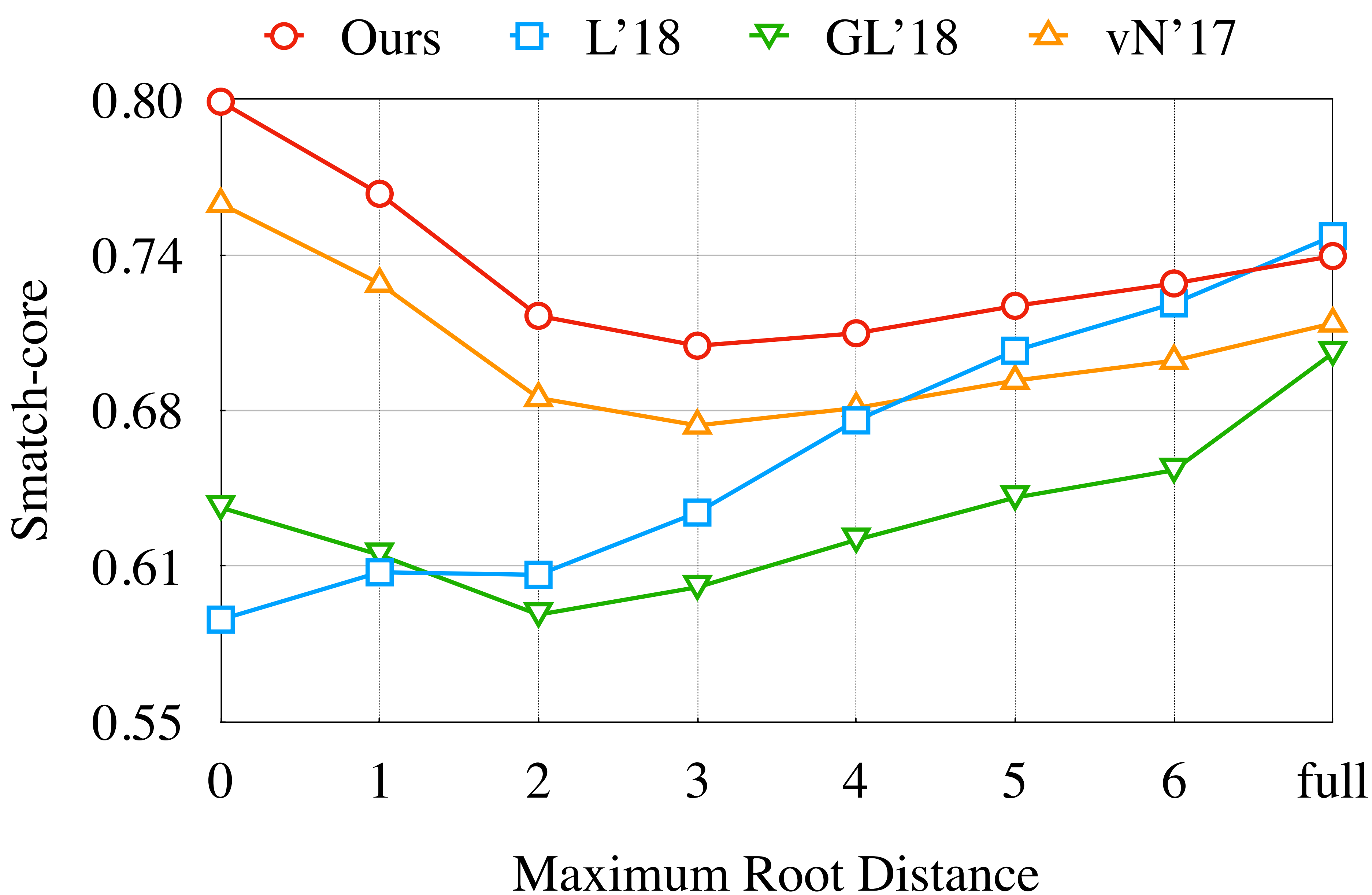
- Seq2seq-based: Buys and Blunsom (2017), van Noord and Bos (2017)
- Transition-based: Guo and Lu (2018)
- Graph-based: Lyu and Titov (2018)
- AM algebra: Groschwitz et al. (2018)

Results and Analysis

Model	Graph Re-ca.	Smatch(%)			RA(%)	CM(%)
		weighted	core	ordinary		
Buys and Blunsom (2017)	No	-	-	61.9	-	-
van Noord and Bos (2017) + 100K	No	68.8	67.6	71.0	75.8	10.2
Guo and Lu (2018)	Yes	63.5	62.3	69.8	63.6	9.4
Lyu and Titov (2018)	Yes	66.6	67.1	**74.4**	59.1	10.2
Groschwitz et al. (2018)	Yes	-	-	71.0	-	-
Ours	No	**71.3**	**70.2**	73.2	**76.9**	**11.6**

- In terms of parser's quality on capturing core semantics, our method significantly outperforms all other methods.
- Competitive results to state-of-the-art even without graph re-categorization (state-of-the-art in the sense that no graph re-ca.).
- RA and CM further confirm the usefulness of a global view and the core semantic first principle

Results and Analysis



- Our method has a clear advantage in capturing the core ideas.

Core Semantic First:
A Top-down Approach for AMR Parsing

Thanks!

Deng Cai and Wai Lam
The Chinese University of Hong Kong

https://github.com/jcyk/AMR-parser
thisisjcykcd@gmail.com