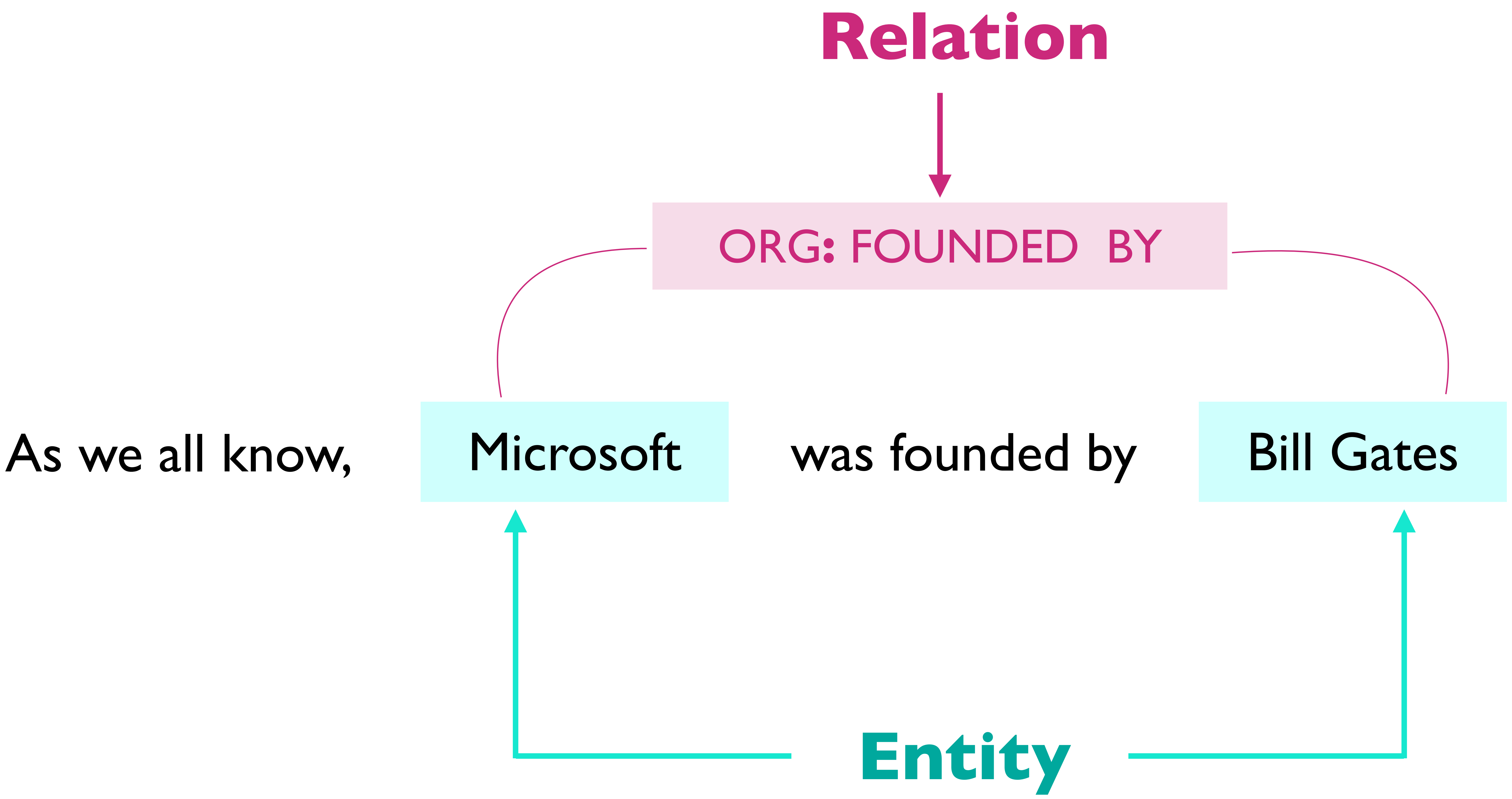
[EMNLP22]Query-based Instance Discrimination Network for Relational Triple Extraction

Zeqi Tan[1], **Yongliang Shen**[1], **Xuming Hu**[2], **Wenqi Zhang**[1], **Xiaoxia Cheng**[1],
Weiming Lu[1]*, **Yueting Zhuang**[1]
[1]Zhejiang University, [2]Tsinghua University

Speaker: 杨晰

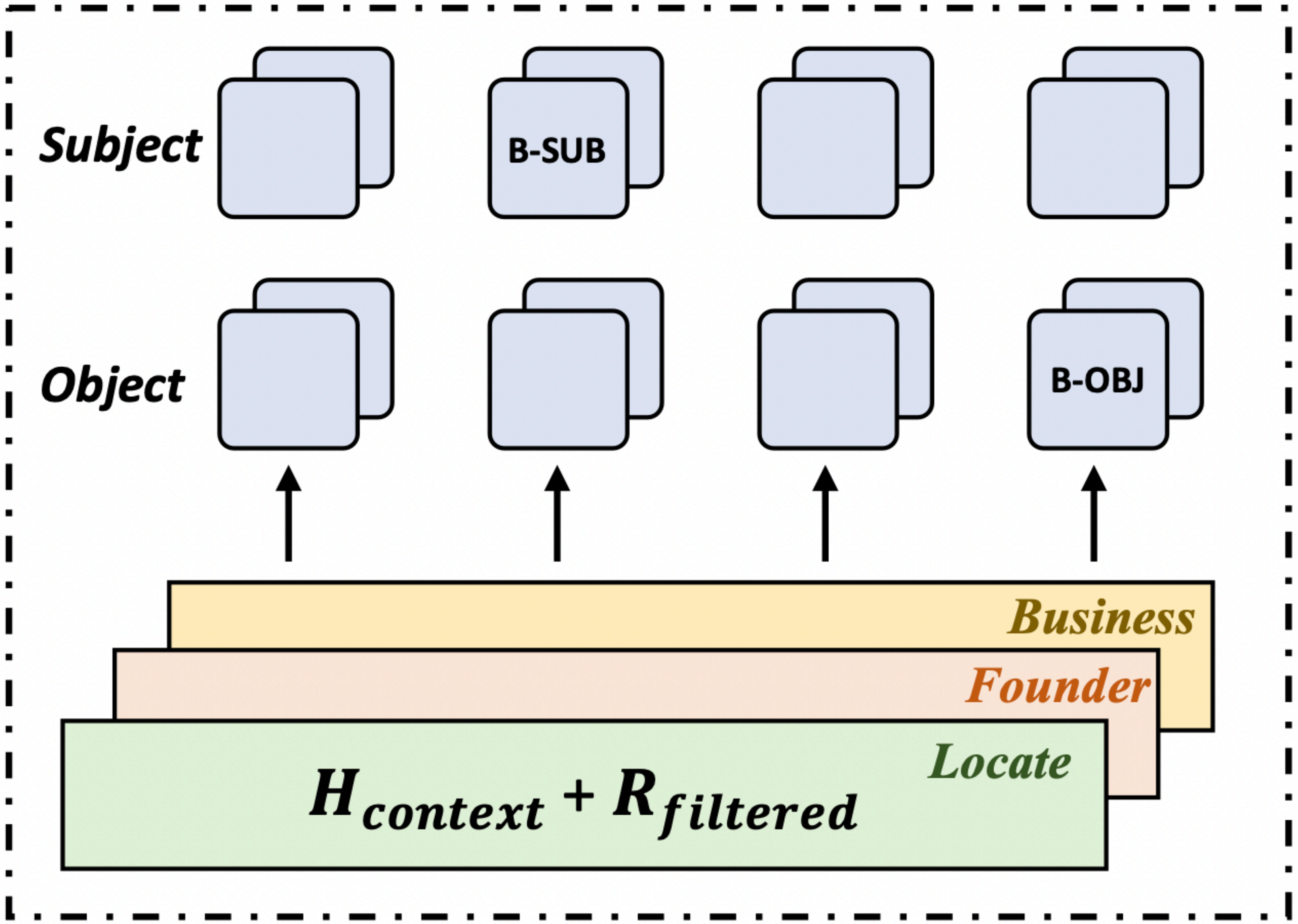
xyang41@stu.ecnu.edu.cn

BG: Relation Extraction



Motivations

PRGC



Relation-specific Tagger

- Error Propagation
- Relation Redundancy
- Processing Triples Independently

Contributions

- Query embeddings: extract all triples in one step
- Instance Discriminator: Contrastive Learning on triple-level representations, for high-order connections

↓

Error Propagation ✅

Relation Redundancy ⬇️

Processing Triples Independently ✅

Task Formulation

Input $X = x_1, x_2, \ldots, x_n$

Entity e (x_i, x_j, t_e)

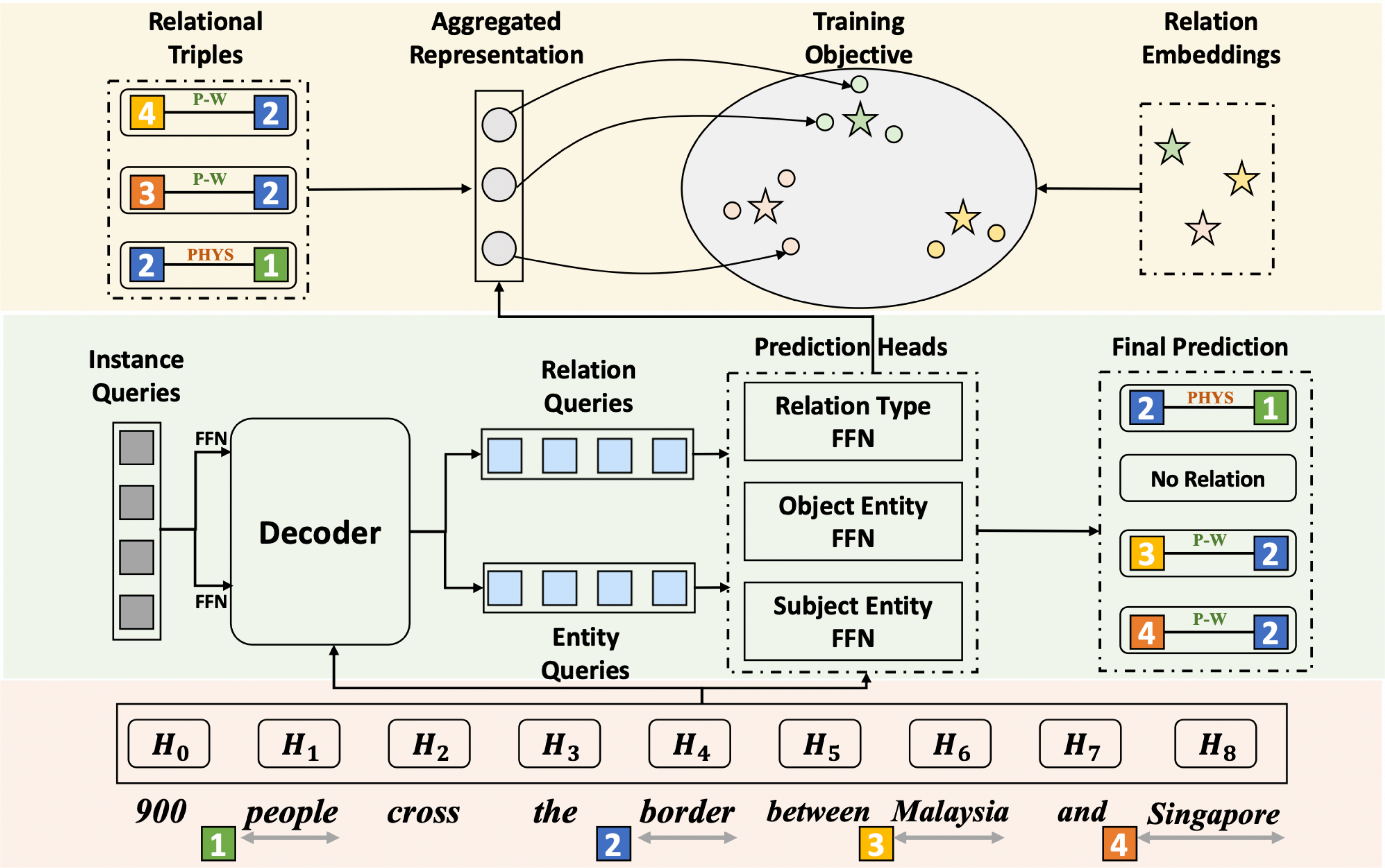
Triple (e_1, e_2, t_r)

Architecture QIDN

III **Instance Discriminator**

II **Triple Prediction**

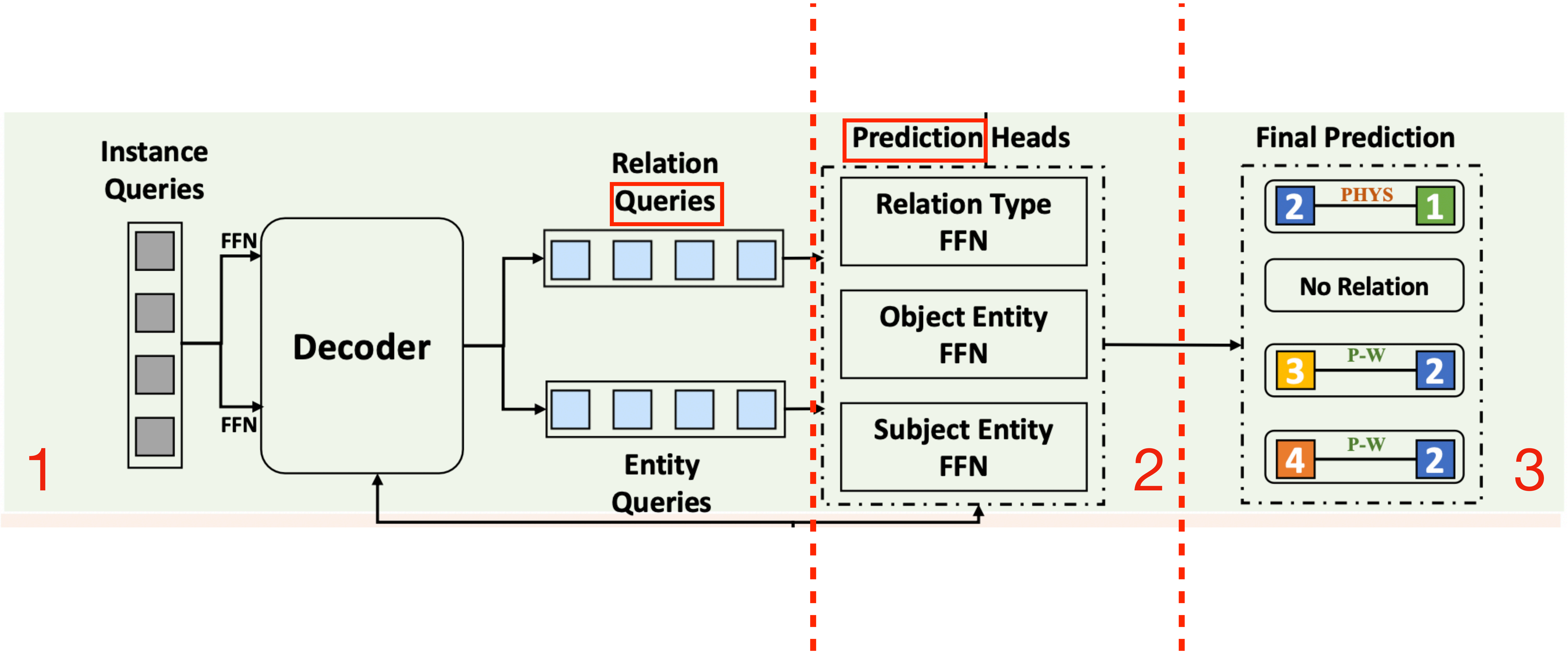
I **Sentence Encoder**



BERT + BiLSTM

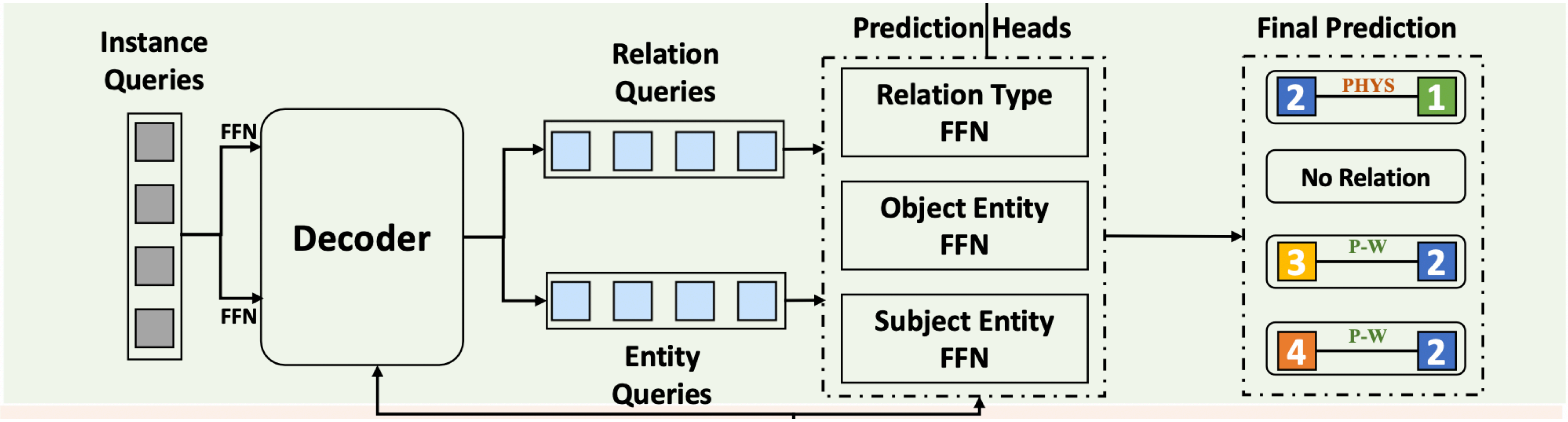
II: Triple Prediction

Triple Prediction

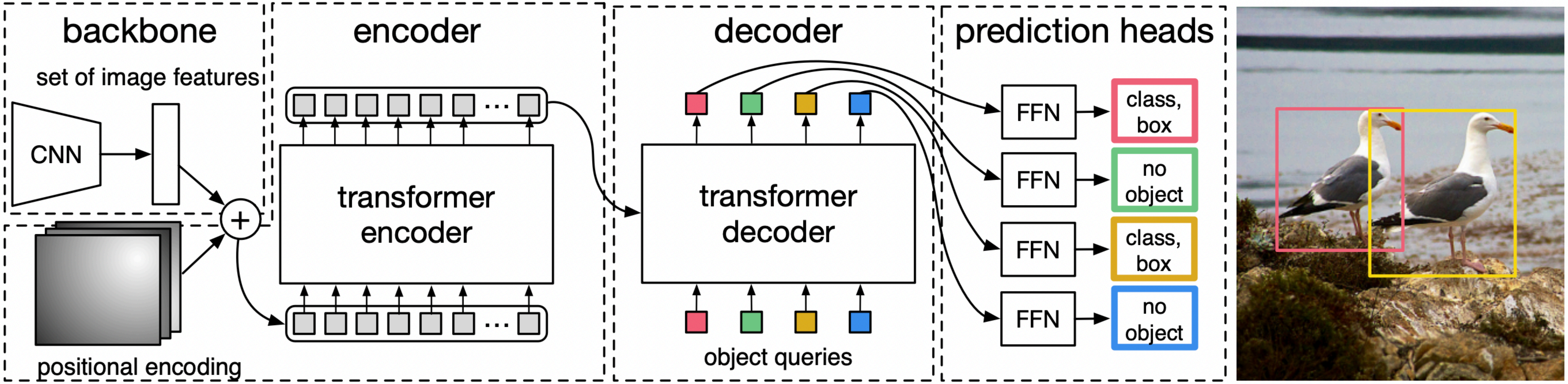


II: Triple Prediction

Triple Prediction

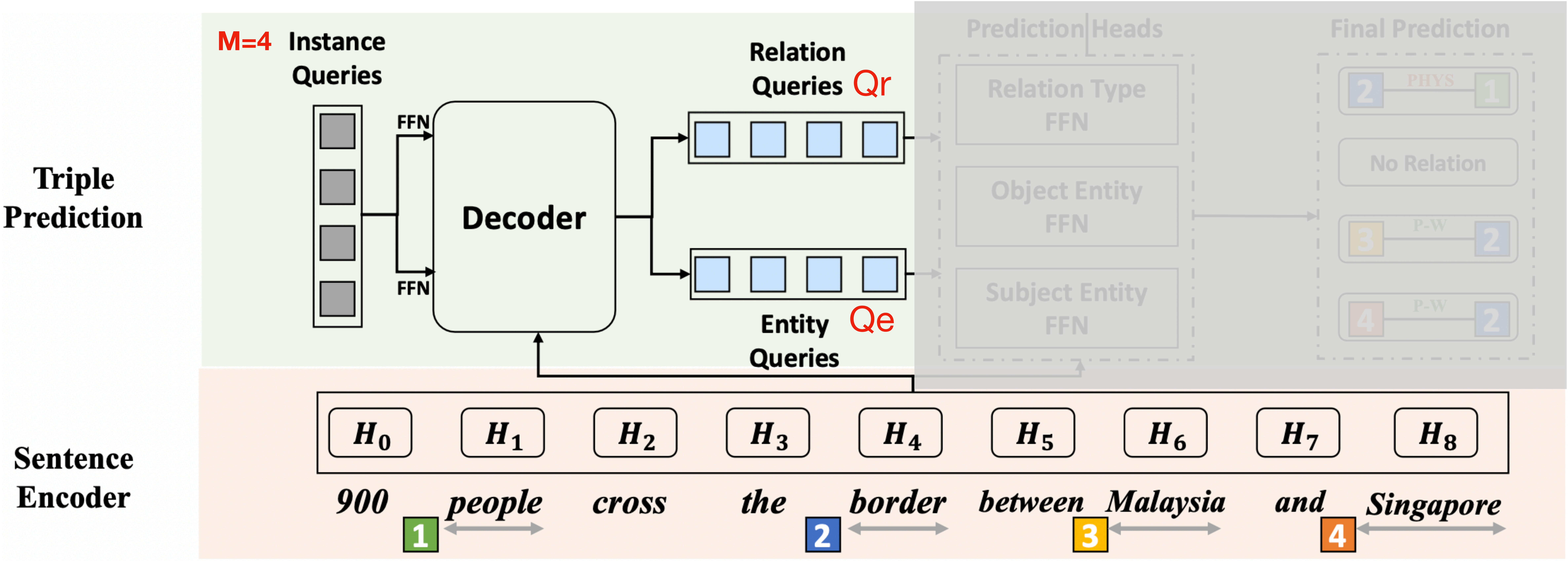


DTER

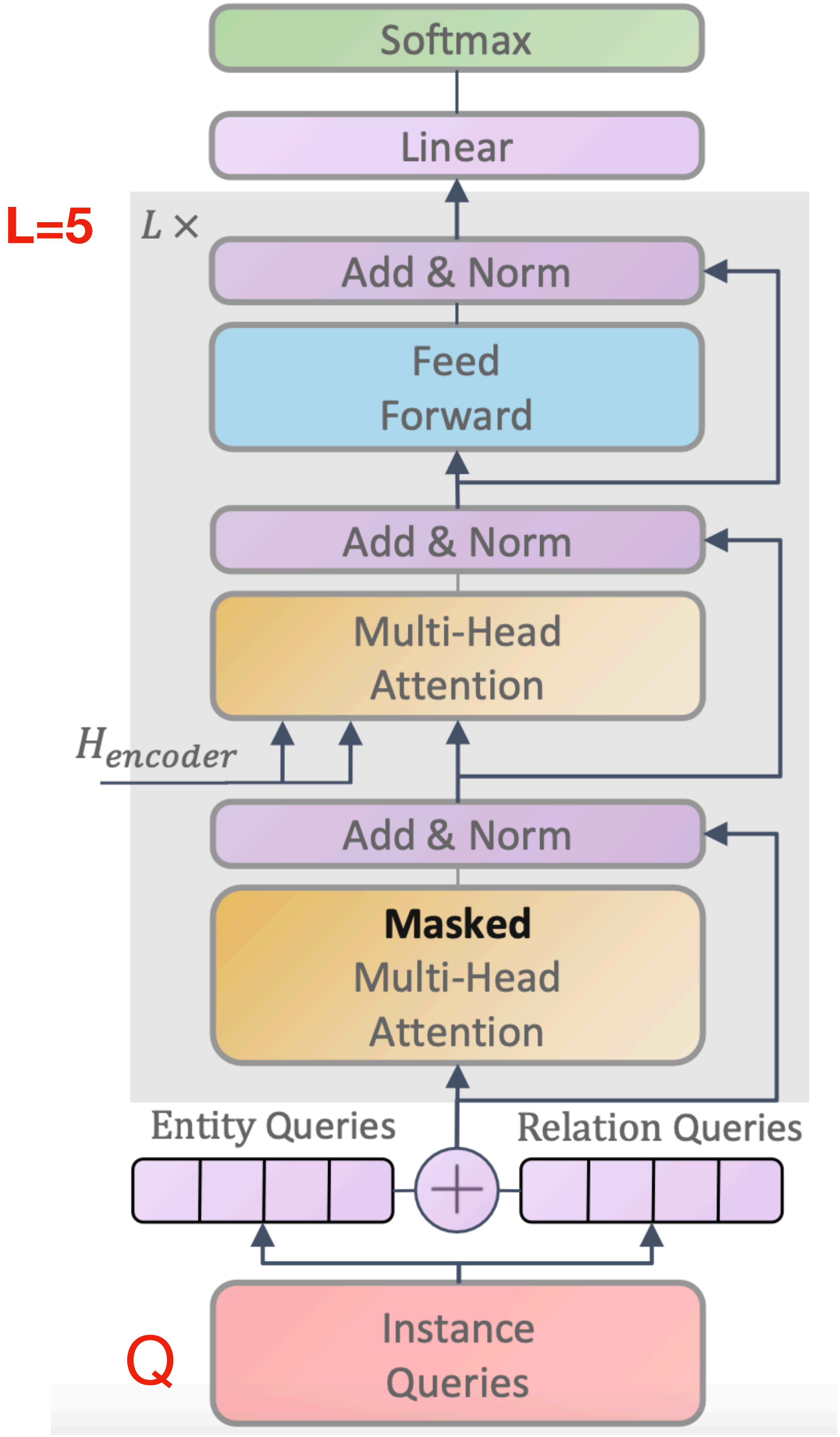


Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-end object detection with transformers. arXiv preprint arXiv:2005.12872.

II: Triple Prediction I



II: Triple Prediction I

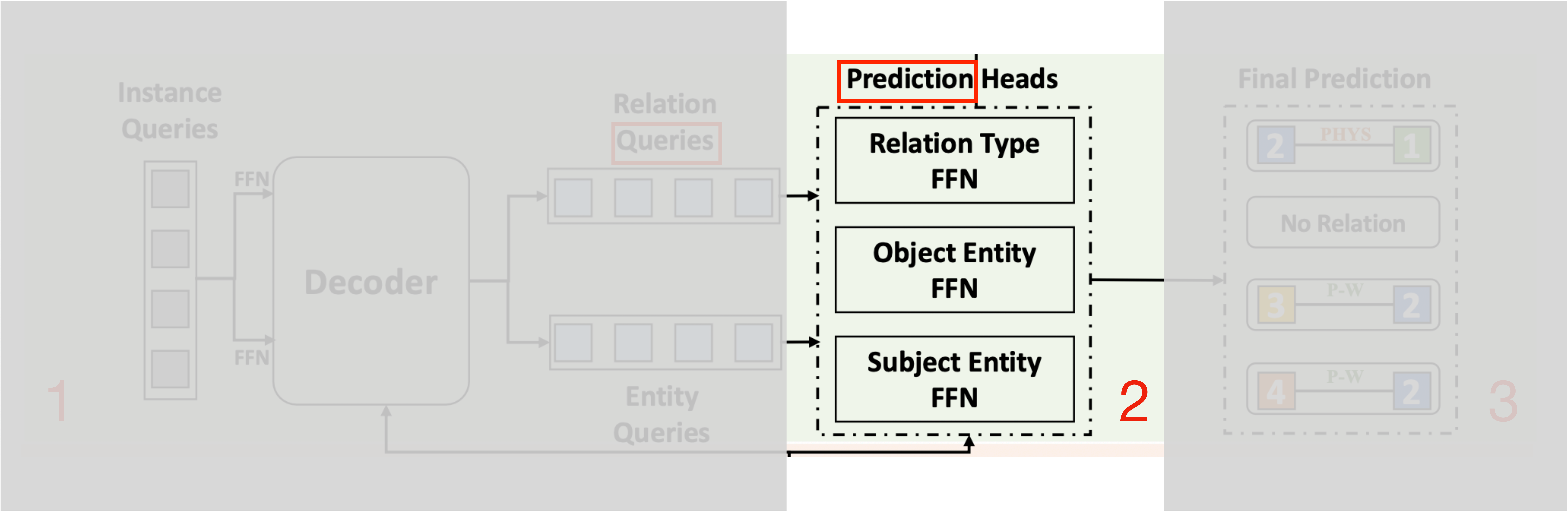


$$[Q_r; Q_e] = \text{Decoder}\left([QW_r; QW_e], H^{span}\right)$$

$$H_{\mathrm{i}}^{span} = \left[H_{\mathrm{start(i)}}; H_{\mathrm{end(i)}}; \phi\left(s_{\mathrm{i}}\right)\right]$$

II: Triple Prediction2

Triple Prediction



II: Triple Prediction2

Relation Head

$$P_{ic}^{t} = \frac{\exp(\boxed{Q_r^i} W_t^c + b_t^c)}{\sum_{c'}^{|\mathcal{Y}_r|} \exp(Q_r^i W_t^{c'} + b_t^{c'})}$$

c: relation type

Entity Head

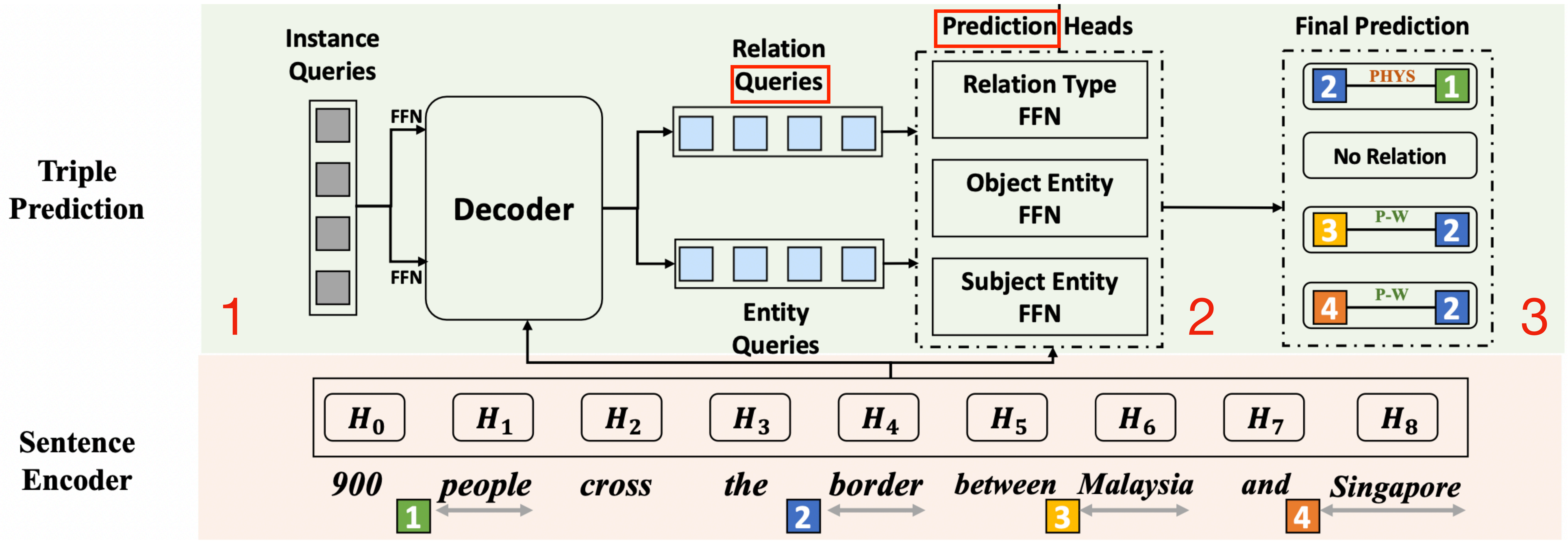
$$E_\delta = \boxed{Q_e} W_\delta, H_s = \boxed{H} W_s$$

$$\delta \in \mathcal{C} = \{l_{sub}, r_{sub}, l_{obj}, r_{obj}\}$$

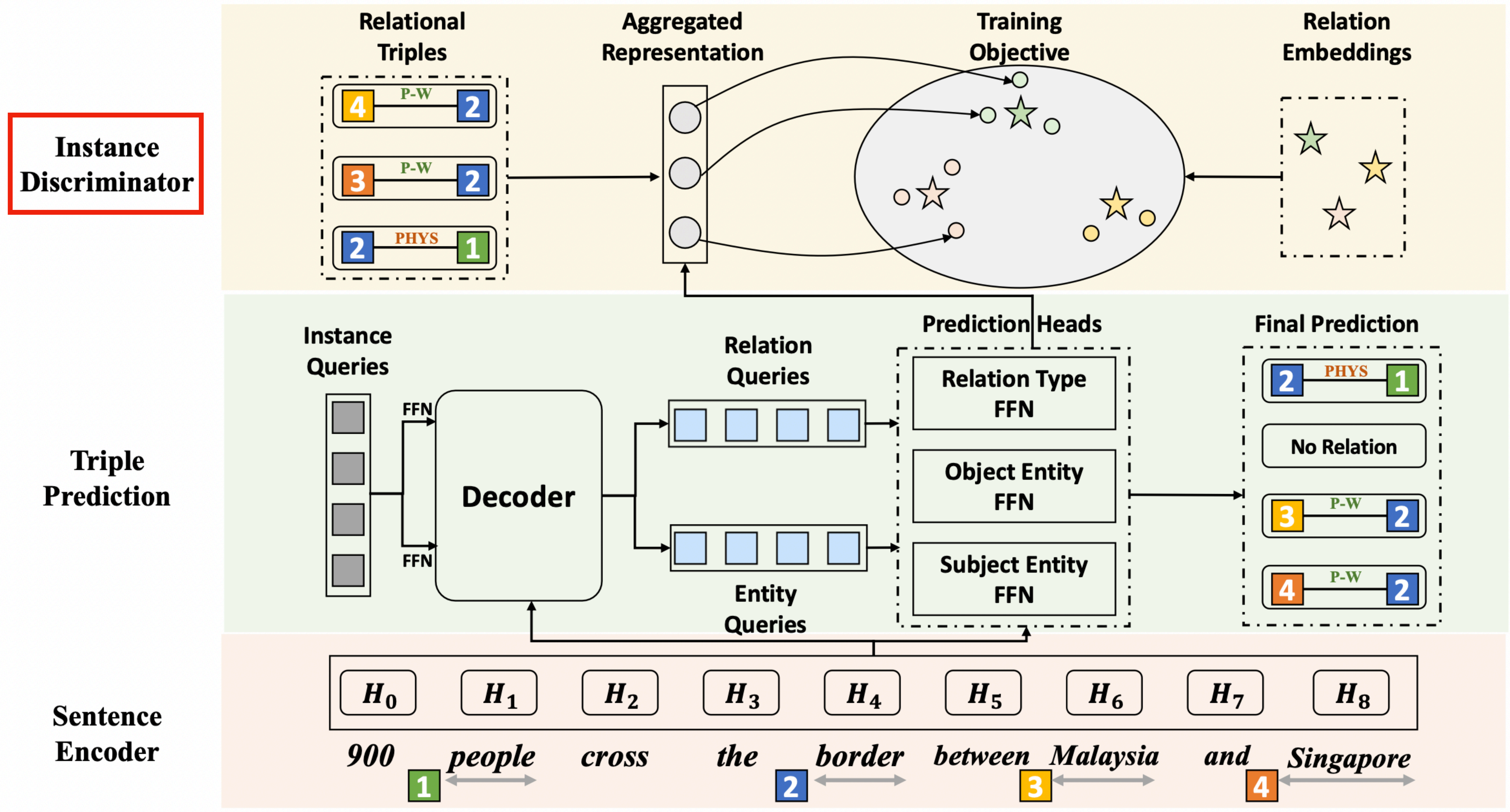
$$P_{ij}^{\delta} = \frac{\exp S\left(\boxed{E_\delta^i}, \boxed{H_s^j}\right)}{\sum_{j'}^{n} \exp S\left(E_\delta^i, H_s^{j'}\right)}$$

$$S\left(\mathbf{v}_i, \mathbf{v}_j\right) = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \cdot \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}$$

II: Triple Prediction3



III: Instance Discriminator



III: Instance Discriminator

Triple instance reprs

$$\mathbf{v} = Q_r W + \sum_{\delta \in \mathcal{C}} E_\delta \qquad \delta \in \mathcal{C} = \{l_{sub}, r_{sub}, l_{obj}, r_{obj}\}$$

Contrastive Learning

instance-instance:

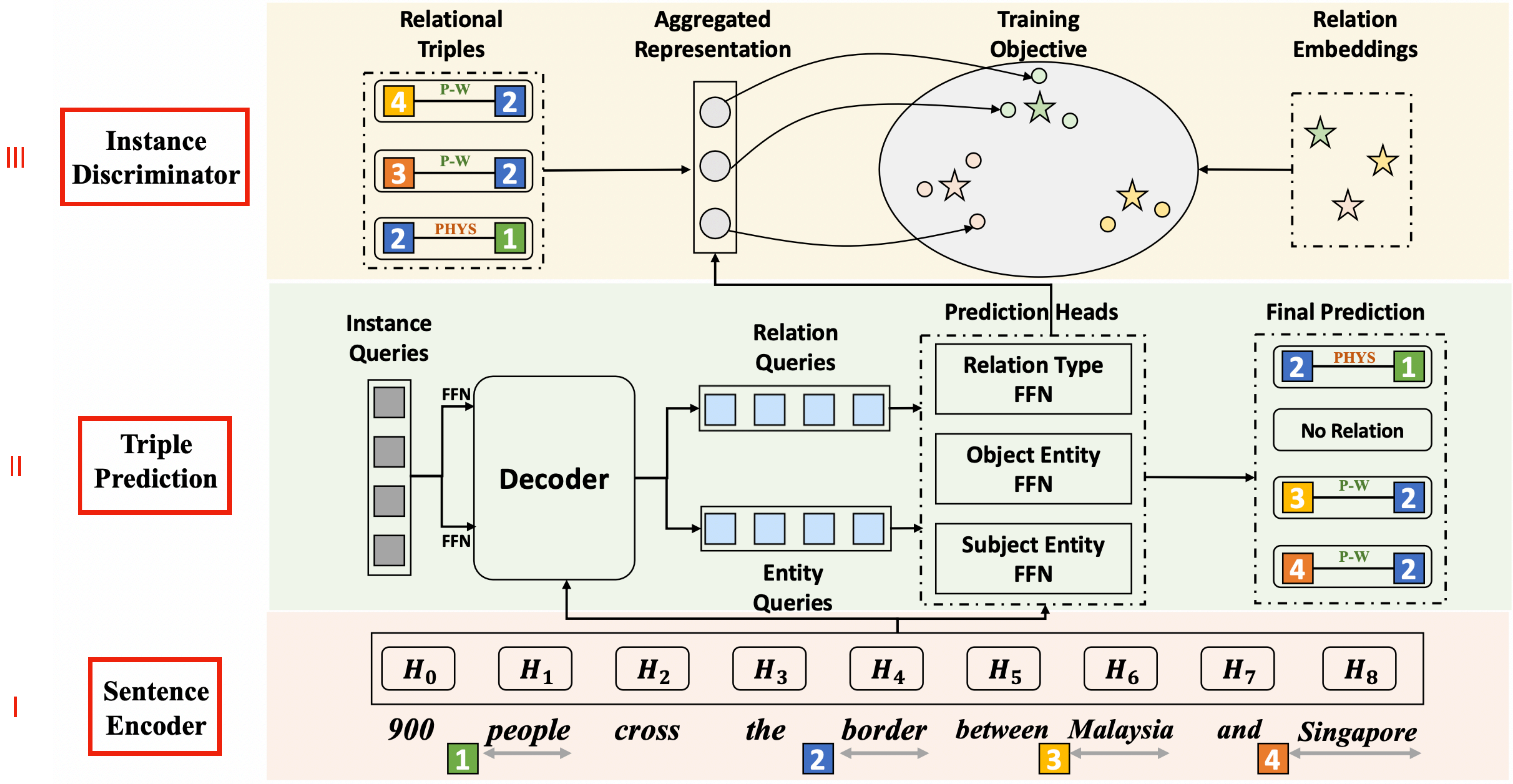
$$\mathcal{L}_{\text{ins}} = -\sum_{c} \sum_{i,j} \log \frac{\exp S\left(\mathbf{v}_i^c, \mathbf{v}_j^c\right)}{\sum_{c\prime, j'} \exp S\left(\mathbf{v}_i^c, \mathbf{v}_{j'}^{c\prime}\right)}$$

i,j: with c relation type

Instance-type:

$$\mathcal{L}_{\text{cls}} = -\sum_{i,c} \log \frac{\exp S\left(\mathbf{v}_i^c, \mathbf{r}_c\right)}{\sum_{c\prime} \exp S\left(\mathbf{v}_i^c, \mathbf{r}_{c'}\right)}$$

Architecture QIDN



BERT + BiLSTM

Training Loss

$$\mathcal{L}_{\text{tri}} = -\sum_{i=1}^{M}\left(\log P^{t}_{\sigma(i)} + \sum_{\delta\in\mathcal{C}} \log P^{\delta}_{\sigma(i)}\right)$$

optimum matching

$$\mathcal{L} = \mathcal{L}_{tri} + \mathcal{L}_{ins} + \mathcal{L}_{cls}$$

$$\mathcal{L}_{\text{ins}} = -\sum_{c}\sum_{i,j}\log\frac{\exp S\left(\mathbf{v}_i^c, \mathbf{v}_j^c\right)}{\sum_{c\prime,j'}\exp S\left(\mathbf{v}_i^c, \mathbf{v}_{j'}^{c\prime}\right)}$$

$$\mathcal{L}_{\text{cls}} = -\sum_{i,c}\log\frac{\exp S\left(\mathbf{v}_i^c, \mathbf{r}_c\right)}{\sum_{c\prime}\exp S\left(\mathbf{v}_i^c, \mathbf{r}_{c'}\right)}$$

Experiments

Dataset	#Sentences			Relations
	Train	Dev	Test	
NYT	56,195	4,999	5,000	24
WebNLG	5,019	500	703	170
NYT*	56,195	5,000	5,000	24
ACE05	10,051	2,424	2,050	6
SciERC	1,861	275	551	7

Table 1: Dataset statistics.

Experiments

Model	NYT			WebNLG			NYT*		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
GraphRel (Fu et al., 2019)	63.9	60.0	61.9	44.7	41.1	42.9	-	-	-
RSAN (Yuan et al., 2020)	-	-	-	-	-	-	85.7	83.6	84.6
MHSA (Liu et al., 2020)	88.1	78.5	83.0	89.5	86.0	87.7	-	-	-
CasRel (Wei et al., 2020)	89.7	89.5	89.6	93.4	90.1	91.8	-	-	-
TPLinker (Wang et al., 2020b)	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0
SPN (Sui et al., 2020)	93.3	91.7	92.5	93.1	93.6	93.4	92.5	92.2	92.3
CGT (Ye et al., 2021)	**94.7**	84.2	89.1	92.9	75.6	83.4	-	-	-
CasDE (Ma et al., 2021)	90.2	90.9	90.5	90.3	91.5	90.9	89.9	91.4	90.6
RIFRE (Zhao et al., 2021)	93.6	90.5	92.0	93.3	92.0	92.6	-	-	-
PRGC (Zheng et al., 2021)	93.3	91.9	92.6	94.0	92.1	93.0	**93.5**	91.9	92.7
QIDN	93.4	**92.6**	**93.0**	**94.1**	**93.7**	**93.9**	93.3	**92.5**	**92.9**

Table 2: Precision(%), Recall (%) and F1-score (%) of our method and baselines. Note that the sentence encoders adopted in GraphRel, RSAN and MHSA are LSTM networks, while other baselines employ the BERT model.

Experiments

Model	NER	RE
ACE05		
Structured Perceptron (Li and Ji, 2014)	80.8	49.5
SPTree (Miwa and Bansal, 2016)	83.4	55.6
Multi-turn QA (Li et al., 2019) †	84.8	60.2
Table-Sequence (Wang and Lu, 2020) ‡	89.5	64.3
Trigger-Sense (Shen et al., 2021b) †	87.6	62.8
PURE (Zhong and Chen, 2021) ‡	89.7	65.6
UniRE (Wang et al., 2021) ‡	90.2	66.0
PFN (Yan et al., 2021) ‡	89.0	66.8
QIDN ‡	**90.5**	**68.2**
SciERC		
SPE (Wang et al., 2020a) §	68.0	34.6
PURE (Zhong and Chen, 2021) §	66.6	35.6
UniRE (Wang et al., 2021) §	68.4	36.9
PFN (Yan et al., 2021) §	66.8	38.4
QIDN §	**69.8**	**39.5**

Table 3: The overall performances of our method on ACE05 and SciERC. †, ‡ and § denotes the use of BERT, ALBERT and SciBERT(Devlin et al., 2019; Lan et al., 2020; Beltagy et al., 2019) pre-trained language models.

Experiments

Model	Encoder	Rep Type	ACE05 Ent	ACE05 Rel	ACE05 Rel+	ACE04 Ent	ACE04 Rel	ACE04 Rel+	SciERC Ent	SciERC Rel	SciERC Rel+
Li and Ji (2014)	-	-	80.8	52.1	49.5	79.7	48.3	45.3	-	-	-
SPtree (Miwa and Bansal, 2016)	LSTM	T	83.4	-	55.6	81.8	-	48.4	-	-	-
DYGIE (Luan et al., 2019)◇	ELMo	T	88.4	63.2	-	87.4	59.7	-	65.2	41.6	-
Multi-turn QA (Li et al., 2019)	BERT_L	-	84.8	-	60.2	83.6	-	49.4	-	-	-
OneIE (Lin et al., 2020)	BERT_L	T	88.8	67.5	-	-	-	-	-	-	-
DYGIE++ (Wadden et al., 2019)◇	BERT_B/SciBERT	T	88.6	63.4	-	-	-	-	-	-	-
TriMF (Shen et al., 2021)◇	BERT_B/SciBERT	T	87.6	66.5	62.8	-	-	-	**70.2**	52.4	-
UniRE (Wang et al., 2021d)◇	BERT_B/SciBERT	T	88.8	-	64.3	87.7	-	60.0	68.4	-	36.9
PURE-F (Zhong and Chen, 2021)◇	BERT_B/SciBERT	S	**90.1**	67.7	64.8	**89.2**	63.9	60.1	68.9	50.1	36.8
PURE-A (Zhong and Chen, 2021)◇	BERT_B/SciBERT	L	-	66.5	-	-	-	-	-	48.1	-
SOTA PL-Marker (Our Model)◇	BERT_B/SciBERT	S&L	89.8	**69.0**	**66.5**	88.8	**66.7**	**62.6**	69.9	**53.2**	**41.6**
TableSeq (Wang and Lu, 2020)	ALB_XXL	T	89.5	67.6	64.3	88.6	63.3	59.6	-	-	-
UniRE (Wang et al., 2021d)◇	ALB_XXL	T	90.2	-	66.0	89.5	-	63.0	-	-	-
PURE-F (Zhong and Chen, 2021)◇	ALB_XXL	S	90.9	69.4	67.0	90.3	66.1	62.2	-	-	-
PL-Marker (Our Model)◇	ALB_XXL	S&L	**91.1**	**73.0**	**71.1**	**90.4**	**69.7**	**66.5**	-	-	-

Experiments: Analysis on Complex Scenarios

Model	NYT									WebNLG								
	Normal	EPO	SEO	SOO	N=1	N=2	N=3	N=4	N≥5	Normal	EPO	SEO	SOO	N=1	N=2	N=3	N=4	N≥5
CasRel	87.3	92.0	91.4	77.0§	88.2	90.3	91.9	94.2	83.7	89.4	94.7	92.2	90.4§	89.3	90.8	94.2	92.4	90.9
TPLinker	90.1	94.0	93.4	90.1§	90.0	92.8	93.1	96.1	90.0	87.9	95.3	92.5	86.0§	88.0	90.1	94.6	93.3	91.6
SPN	90.8	94.1	94.0	-	90.9	93.4	**94.2**	95.5	90.6	-	-	-	-	89.5	91.3	**96.4**	94.7	93.8
PRGC	**91.0**	94.5	94.0	81.8	**91.1**	93.0	93.5	95.5	93.0	90.4	**95.9**	93.6	94.6	89.9	91.6	95.0	94.8	92.8
QIDN	**91.2**	**94.9**	**94.8**	**90.7**	90.6	**93.6**	94.1	**95.8**	**94.3**	**91.5**	95.4	**94.8**	**94.9**	**91.2**	**92.8**	96.1	**95.4**	**94.2**

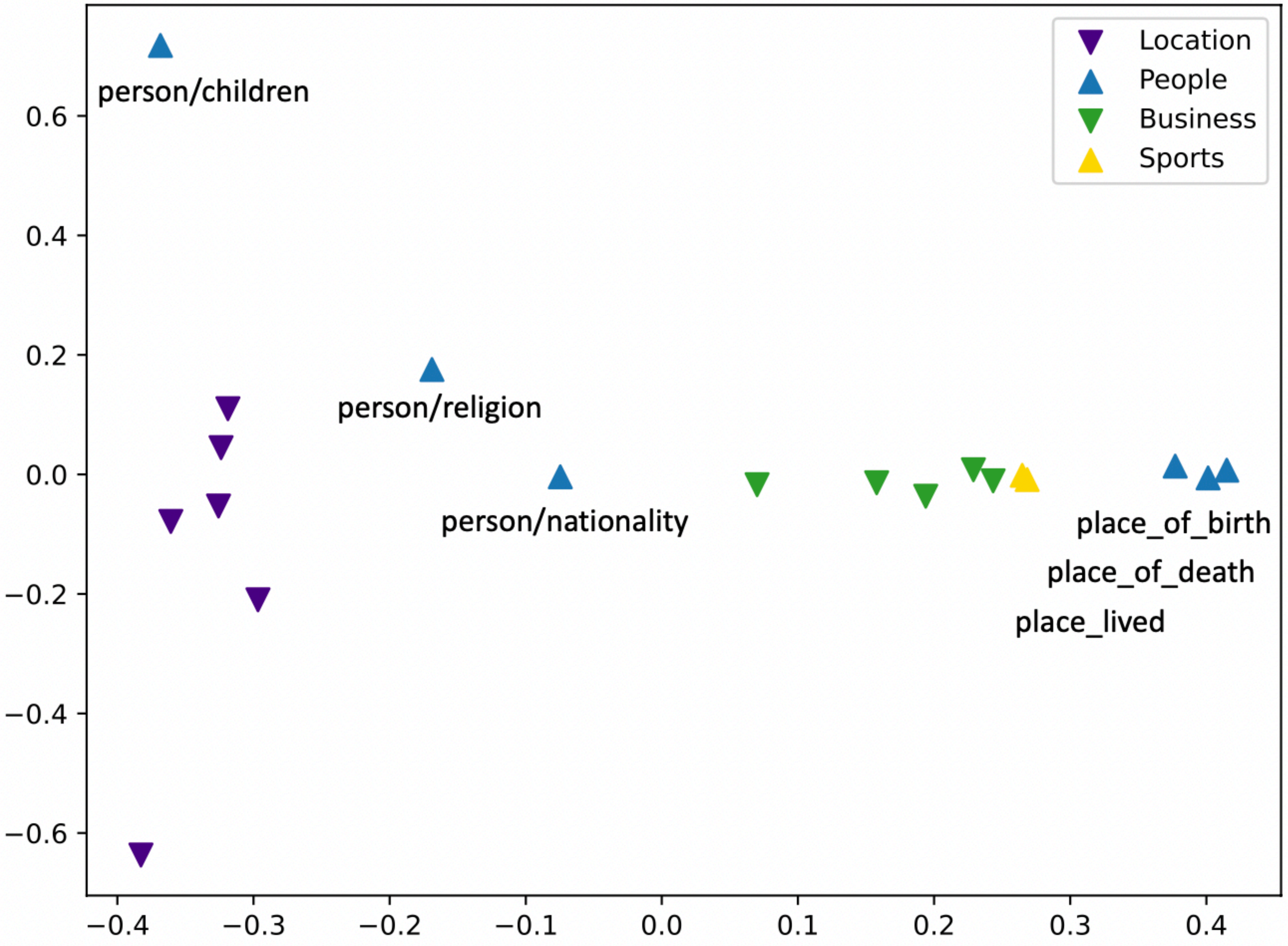
Table 4: F1-measure (%) on sentences with different overlapping patterns and different triple numbers. § marks the results reported by (Zheng et al., 2021).

more robust than baselines

Experiments: Ablation Study

Model	NYT			WebNLG		
	P	R	F	P	R	F
Default	**93.4**	**92.6**	**93.0**	**94.1**	**93.7**	**93.9**
w/o H_{span}	92.9	92.6	92.8	93.5	93.0	93.3
w/o Q_e, Q_r	93.2	92.2	92.7	93.4	92.8	93.1
w/o $\mathcal{L}_{\text{ins}}$	92.0	92.5	92.2	93.0	92.8	92.9
w/o $\mathcal{L}_{\text{cls}}$	92.7	92.2	92.4	93.7	92.5	93.1
Contrastive Learning: w/o $\mathcal{L}_{\text{ins}}, \mathcal{L}_{\text{cls}}$	91.1	92.6	91.8	93.3	91.8	92.5

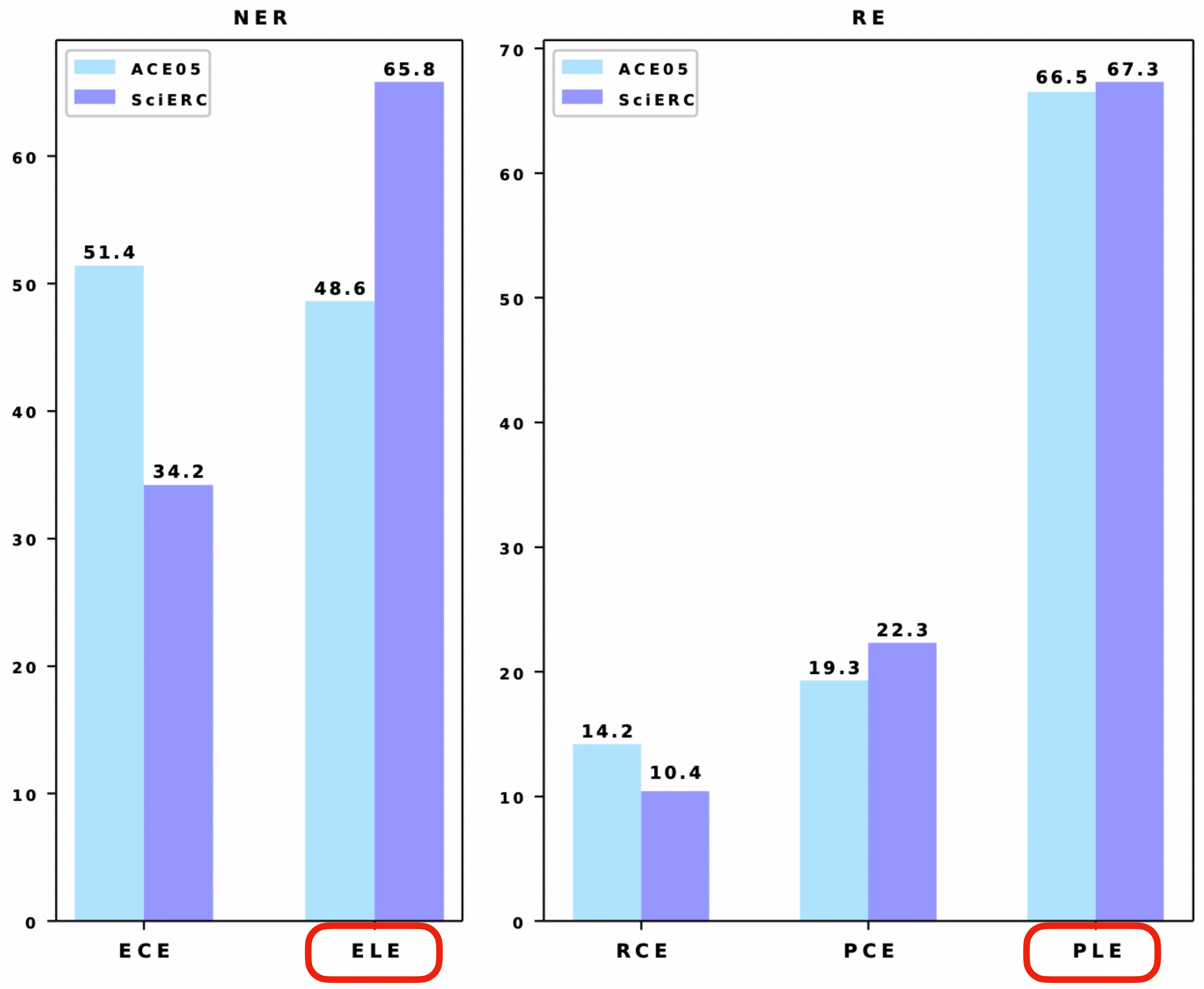
Experiments: Topology of Relations



learn the high-order connections between different relation types

Figure 3: Visualization of relations on NYT dataset. For simplicity, we only label the fine-grained categories under "People".

Experiments: Error Analysis



E: Entity

R: Relation

C: Classification

L: Location

P: Entity-Pair

Figure 4: The proportions of different errors in NER and RE on the ACE05 and SciERC test sets.

Experiments: Analysis of query branches

Inter-task	NER			RE		
	P	R	F	P	R	F
Default	**68.1**	**71.2**	**69.6**	**39.7**	**38.9**	**39.3**
w/o ent-rel	67.6	70.9	69.2	38.5	37.9	38.2
w/o rel-ent	67.5	70.4	68.9	39.5	38.3	38.9
w/o both	67.2	69.8	68.5	38.3	37.5	37.9

Table 7: The performance with different attention mask settings on the SciERC development set.

Experiments: Case Study

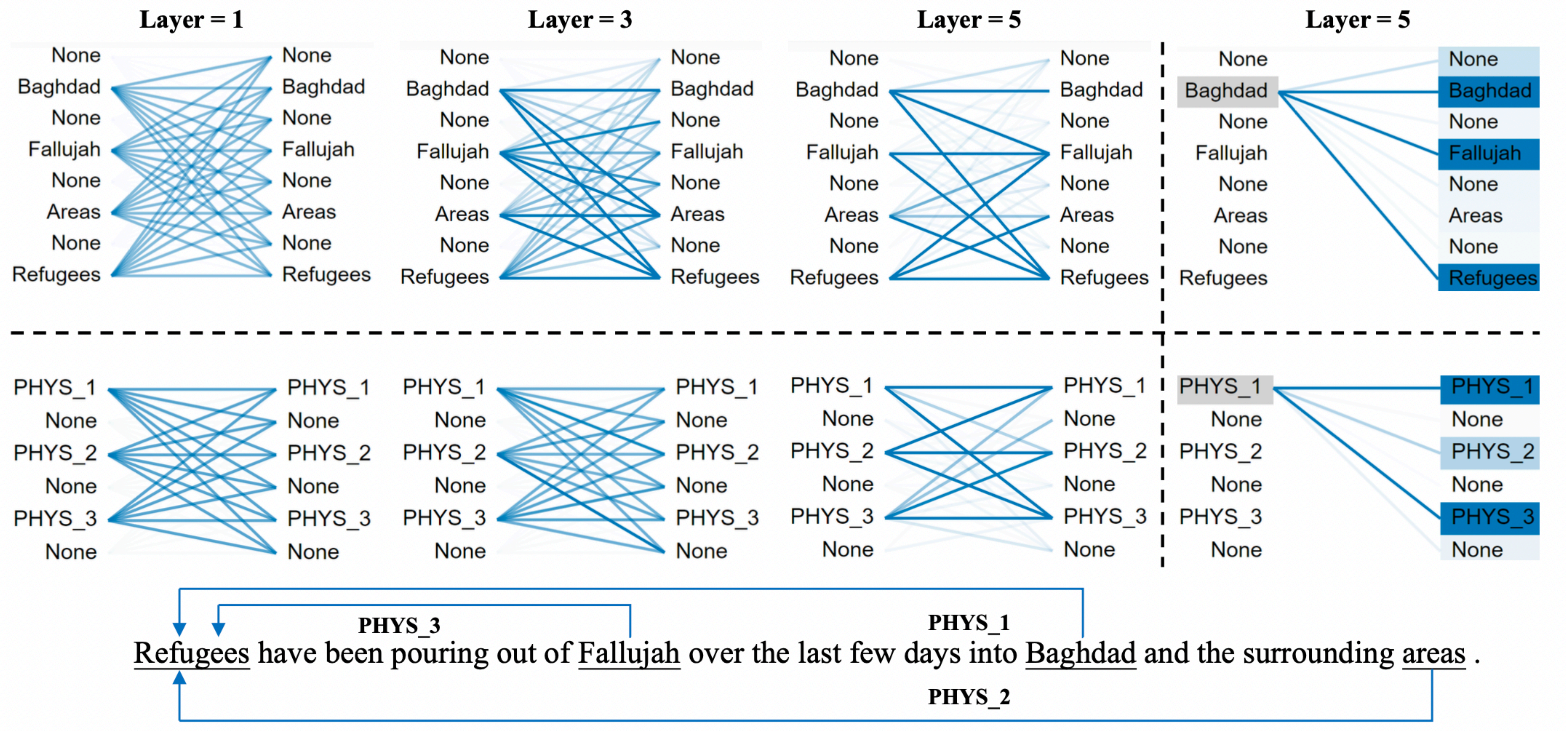


Figure 6: Visualization for the attention weights between entity queries and between relation queries. We randomly select “None” with the same number of gold labels, which indicates a non-entity or a non-relation corresponding to a query, and the attention of “None” is ignored. The sentence is randomly selected from the ACE05 corpus.