AMR Parsing as Graph Prediction with Latent Alignment

Chunchuan Lyu*, Ivan Titov*†

*ILCC, School of Informatics, University of Edinburgh †ILLC, University of Amsterdam

Abstract Meaning Representations

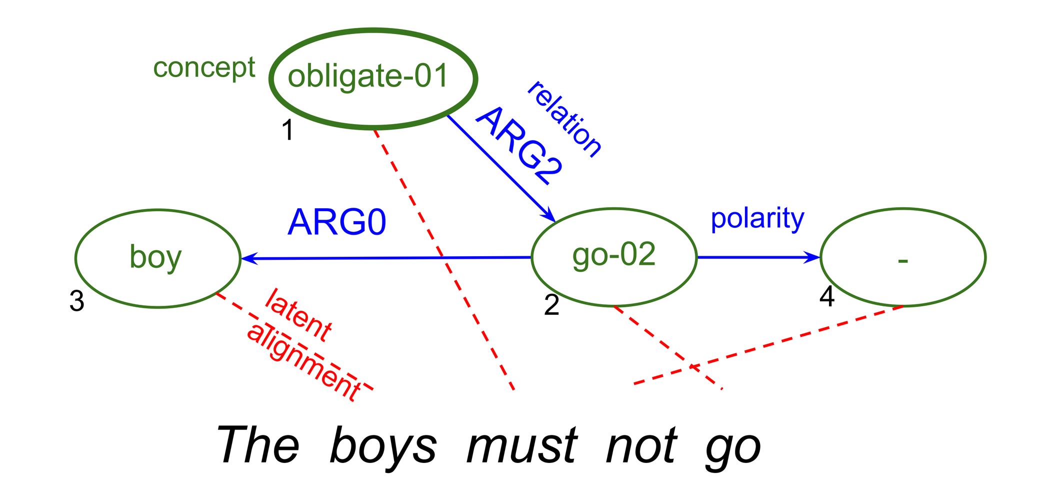


Fig. 1. An example of AMR, the dashed lines denote latent alignments, obligate-01 is the root.

Main Contributions

- Lack of gold alignment -> **AMR parsing with a joint probabilistic model for alignment, concept and relation identification.**
- Seq2seq model could work well for semantic parsing ? -> **our non-autoregressive model achieves the best reported results (+3.4% over previous state of the art).** Sequence tagging does not suffer from exposure bias.

AMR Parsing as Graph Prediction

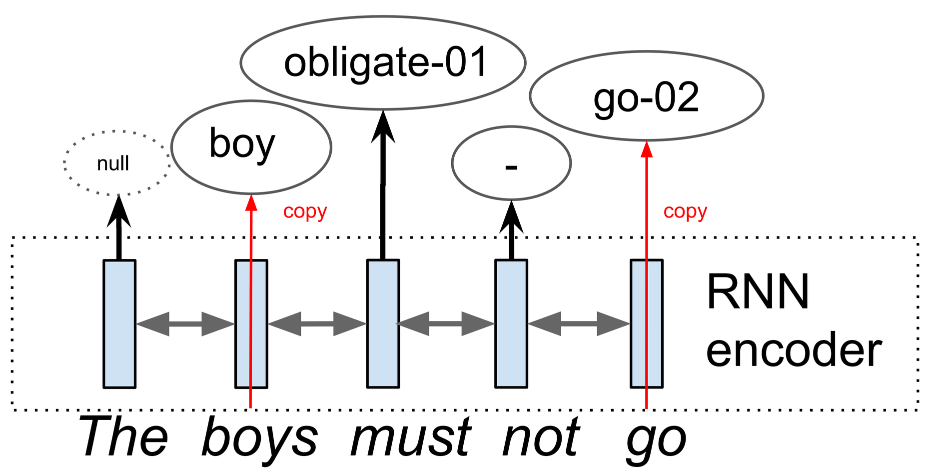


Fig. 2. Concept Identification

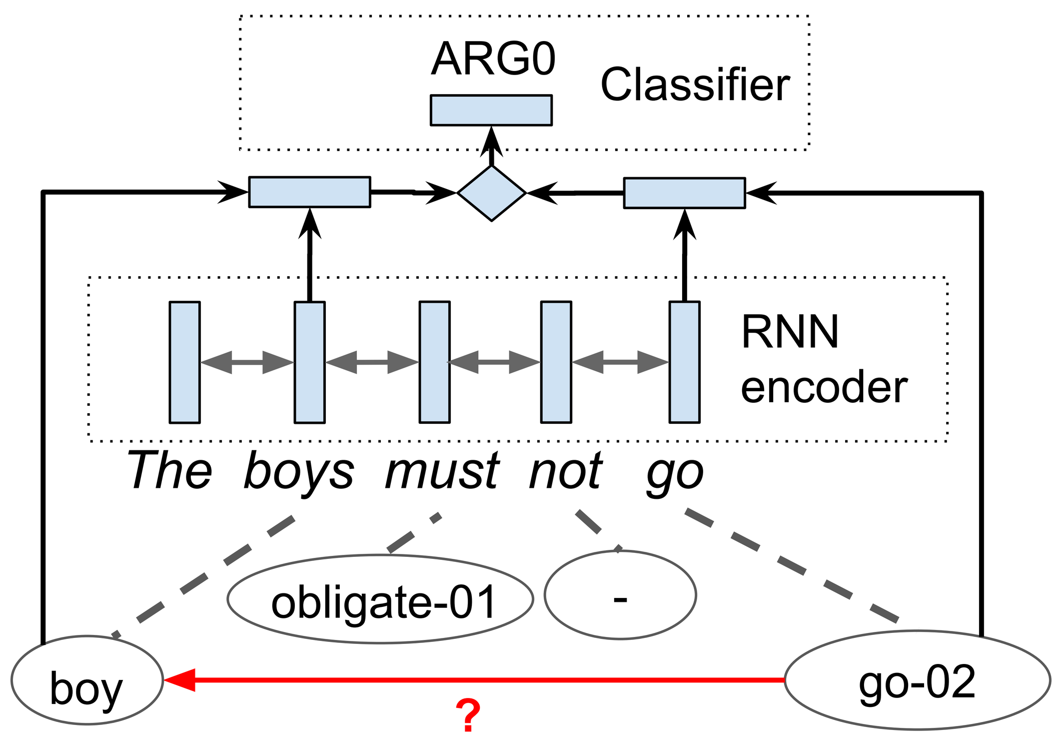


Fig. 3. Relation Identification:

Variational Auto-Encoder

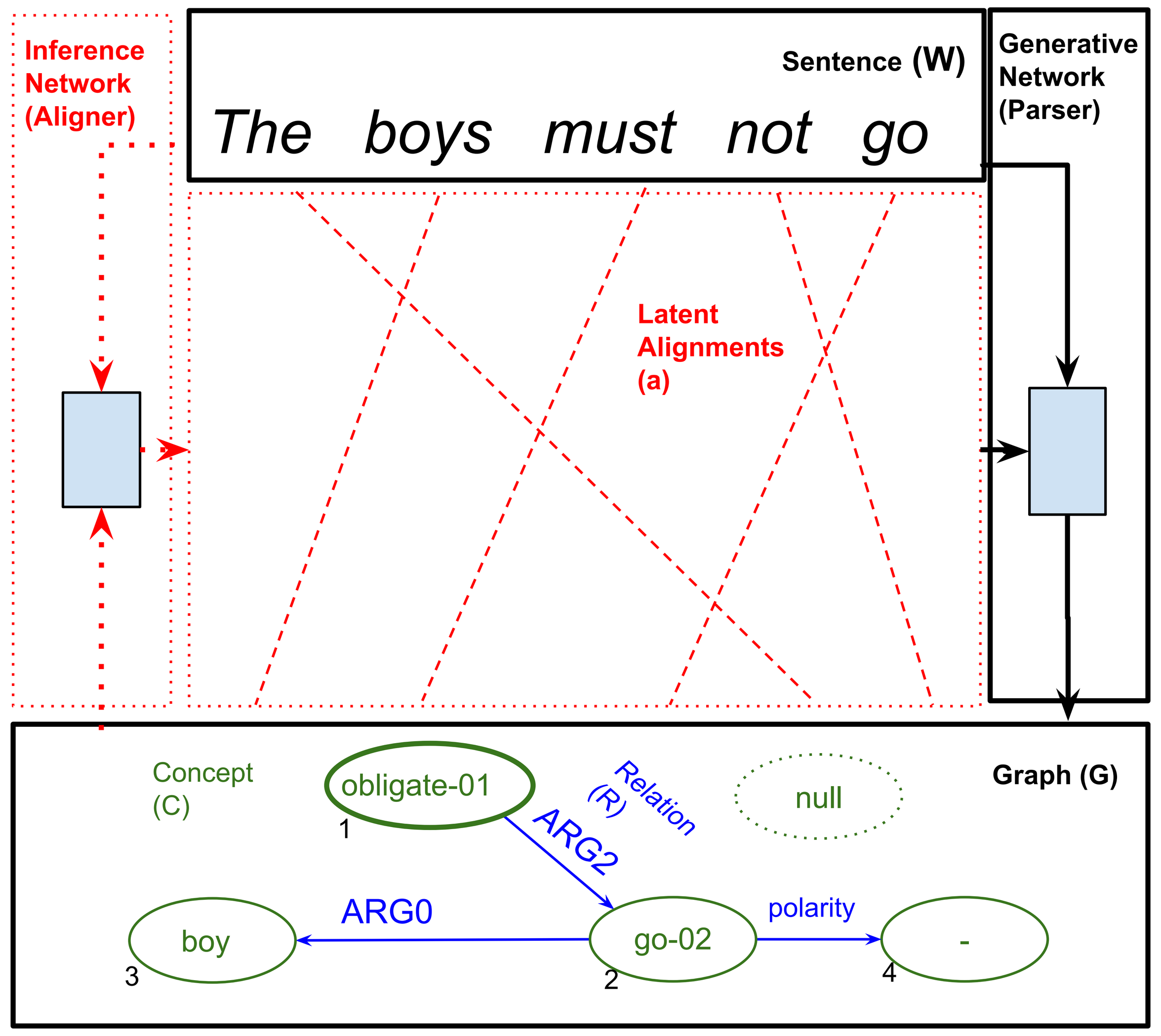


Fig. 4. Joint Training for AMR parsing

Joint Training Objective

$P_{\theta,\phi}(\mathbf{c}, R|\mathbf{w})$ Conditional probability needs marginalizing over bi-jective alignment

$$= \sum_{\mathbf{a} \in Perm} P(\mathbf{a}) P_\theta(\mathbf{c}|\mathbf{a}, \mathbf{w}) P_\phi(R|\mathbf{a}, \mathbf{w}, \mathbf{c})$$

Further conditional independence

$$= \sum_{\mathbf{a} \in Perm} P(\mathbf{a}) \prod_{i=1}^{m} P(c_i|\mathbf{h}_{a_i}) \prod_{i,j=1}^{m} P(r_{ij}|\mathbf{h}_{a_i}, \mathbf{c}_i, \mathbf{h}_{a_j}, \mathbf{c}_j)$$

- marginalization is intractable

-> variational inference

Variational Lower Bound

$\log P_{\theta,\phi}(\mathbf{c}, R|\mathbf{w})$ Log of marginalized probability

$\geq E_Q[\log P_\theta(\mathbf{c}|\mathbf{a}, \mathbf{w}) P_\phi(R|\mathbf{a}, \mathbf{w}, \mathbf{c})]$ Expected likelihood

$-D_{KL}(Q_\psi(\mathbf{a}|\mathbf{c}, R, \mathbf{w}) || P(\mathbf{a})),$ KL regularizer

- Sampling over permutation is still intractable

-> For permutation, Gumbel-Sinkhorn provides relaxed sample from approximate Perturb-and-MAP

(Mena Et all. 2018)

Gumbel-Sinkhorn Relaxation

$\Sigma_{ij} \sim \mathcal{G}(0, 1)$ Generate Gumbel noise for each concept word pair

$\Phi_{ij} = \mathbf{c}_i^T B \mathbf{w}_j$ Compute bilinear score based on RNN encoding

$\hat{\mathbf{a}} = S_t(\Phi, \Sigma)$ Apply Sinkhorn operator to get soft alignment

$S_0(\Phi, \Sigma) = \exp(\Phi + \Sigma)$ Sinkhorn initialize with exponential

$S_t(\Phi, \Sigma) = \mathcal{N}_r(\mathcal{N}_c(S_{t-1}(\Phi, \Sigma)))$ Iteratively normalize across columns and rows

$E_{\Sigma \sim \mathcal{G}(0,1)}[\log P_\theta(c|S_t(\Phi_\psi, \Sigma), \mathbf{w})$
$+ \log P_\phi(R|S_t(\Phi_\psi, \Sigma), \mathbf{w}, \mathbf{c})]$ Feed soft alignment to neural model

$-D_{KL}(\frac{\Phi_\psi + \Sigma}{t} || \frac{\Sigma}{t_0})$ Reparemetrized KL to be computable

- Soft alignment can not be used directly

-> Model relaxation is needed

Model Relaxation

For concept identification model , treat soft alignment as prior

$$\log P_\theta(c_i|\hat{\mathbf{a}}_\mathbf{i}, \mathbf{w}) \approx \log \sum_{k=1}^{n} \hat{\mathbf{a}}_{ik} P_\theta(c_i|a_i = k, \mathbf{w})$$

For relation identification , weight representation with soft alignment

$$\hat{\mathbf{h}_{a_i}} := \sum_{k=1}^{n} \hat{\mathbf{a}}_{ik} \mathbf{h}_k$$

Recategorization

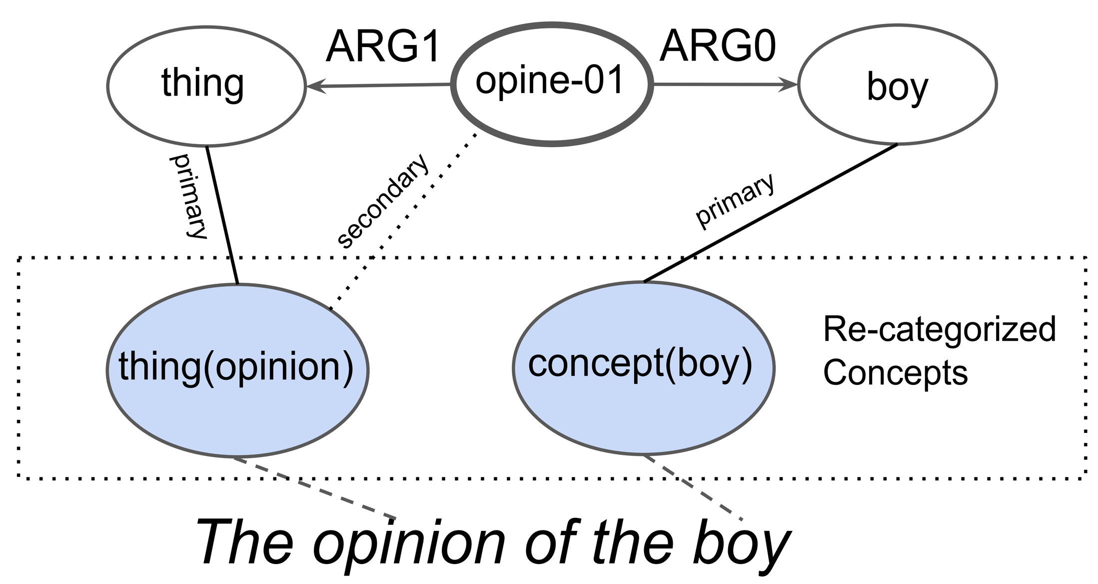
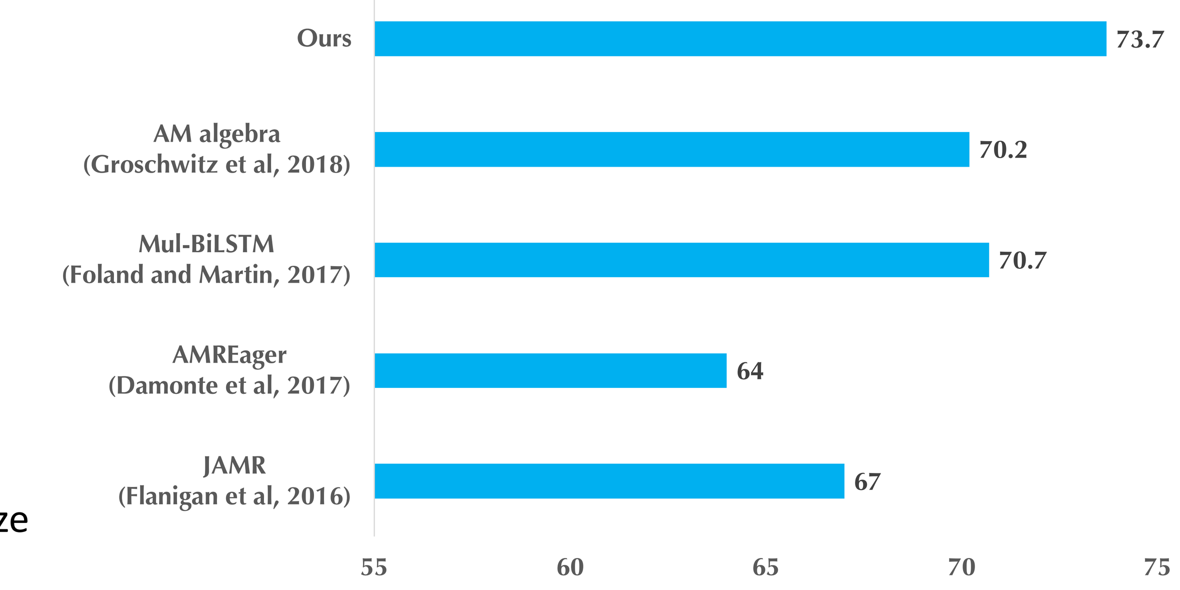
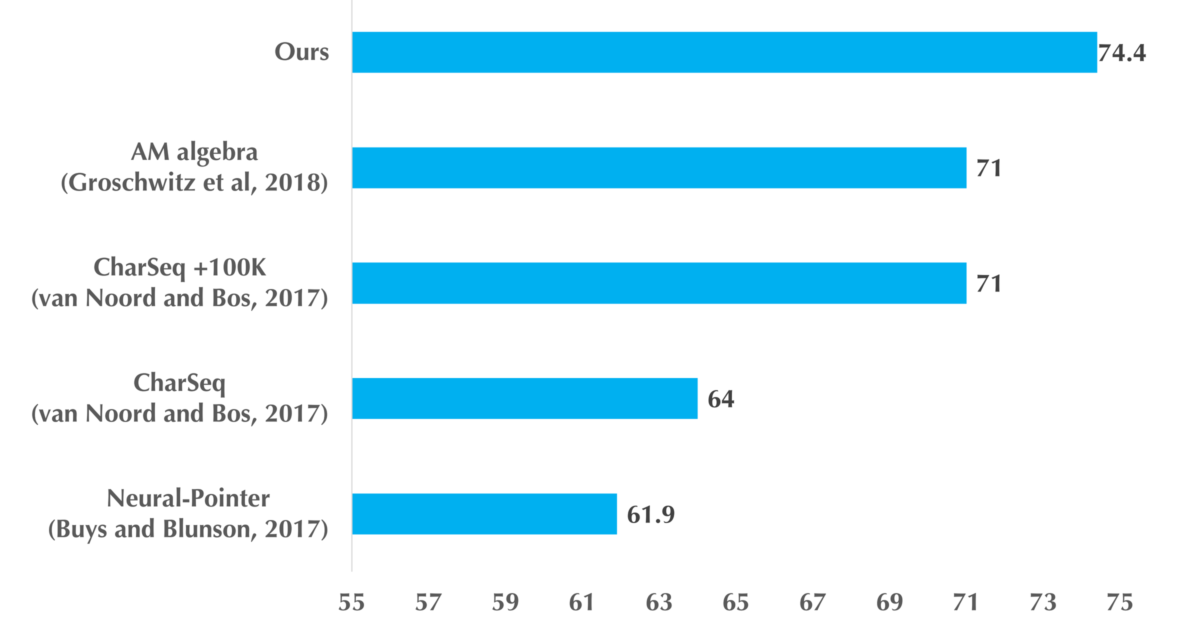


Fig. 5 An example of re-categorized AMR

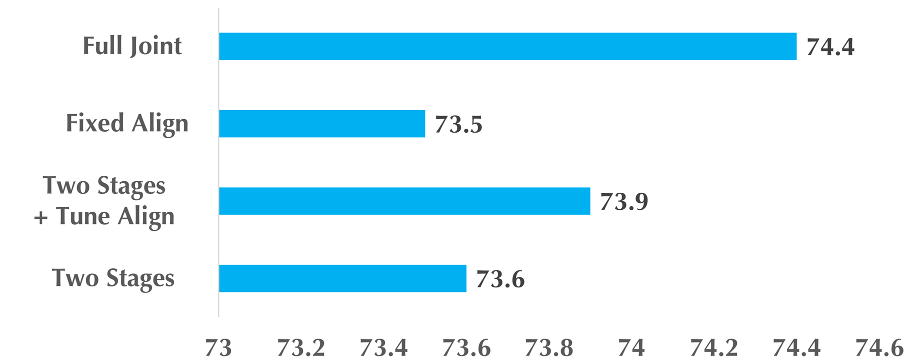
Results





Models	A'17	J'16	Ch'17	AM	Ours
Dataset		R1			R2
Smatch	64	67	71	71	***74.4±0.16***
Unlabeled	69	69	74	74	***77.1±0.10***
No WSD	65	68	72	72	***75.5±0.12***
Concepts	83	83	82	84	***85.9±0.11***
NER	83	79	79	78	***86.0±0.46***
Negations	48	45	***62***	57	58.4±1.32
SRL	56	60	66	64	***69.8±0.24***

Table 1. F1 scores on individual phenomena. A'17 is AMREager, J'16 is JAMR, Ch'17 is CharSeq+100K, AM is AM algebra.



Acknowledgements: We thank Marco Damonte, Shay Cohen, Diego Marcheggiani and Wilker Aziz for helpful discussions as well as anonymous reviewers for their suggestions. The project was supported by the European Research Council (ERC StG BroadSem 678254) and the Dutch National Science Foundation (NWO VIDI 639.022.518).

* Contact: *chunchuan.lv@gmail.com*