# Learning Cross-lingual Distributed Logical Representations for Semantic Parsing

**Yanyan Zou** and Wei Lu
StatNLP Group
Singapore University of Technology and Design
July 18, 2018

# Outline

- ✓ Background & Motivation
- ✓ Method
- ✓ Experiments & Analysis
- ✓ Conclusion

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# Semantic Parsing

Goal: Map natural languages into semantic representations.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# Semantic Parsing

Goal: Map natural languages into semantic representations.

| Natural Language | English: what states have no bordering state ? |
|---|---|

# Semantic Parsing

Goal: Map natural languages into semantic representations.

| | |
|---|---|
| Natural Language | English: what states have no bordering state ? |
| Logical Form | $answer(exclude(state(all), next\_to(state(\mathrm{all}))))$ |

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# Semantic Parsing

Goal: Map natural languages into semantic representations.

| | |
|---|---|
| Natural Language | English: what states have no bordering state ? |
| Semantic Tree | QUERY : *answer* (STATE)<br>STATE: *exclude* (STATE, STATE)<br>STATE : *state* (all)    STATE : *next_to* (STATE)<br>STATE : *state* (all) |
| Logical Form | *answer*(*exclude*(*state*(*all*), *next_to*(*state*(all)))) |

# Joint Representations

Proposed in previous works:

- ✓ Synchronous CFG derivation trees
  Wong and Mooney (2006, 2007)
- ✓ CCG derivation trees
  Zettlemoyer and Collins (2005, 2007)
- ✓ Bayesian tree transducers
  Jones, Goldwater and Johnson (2012)
- ✓ Hybrid Trees
  Lu, Ng, Lee, Zettlemoyer (2008)

 (Lu et al.，2008)

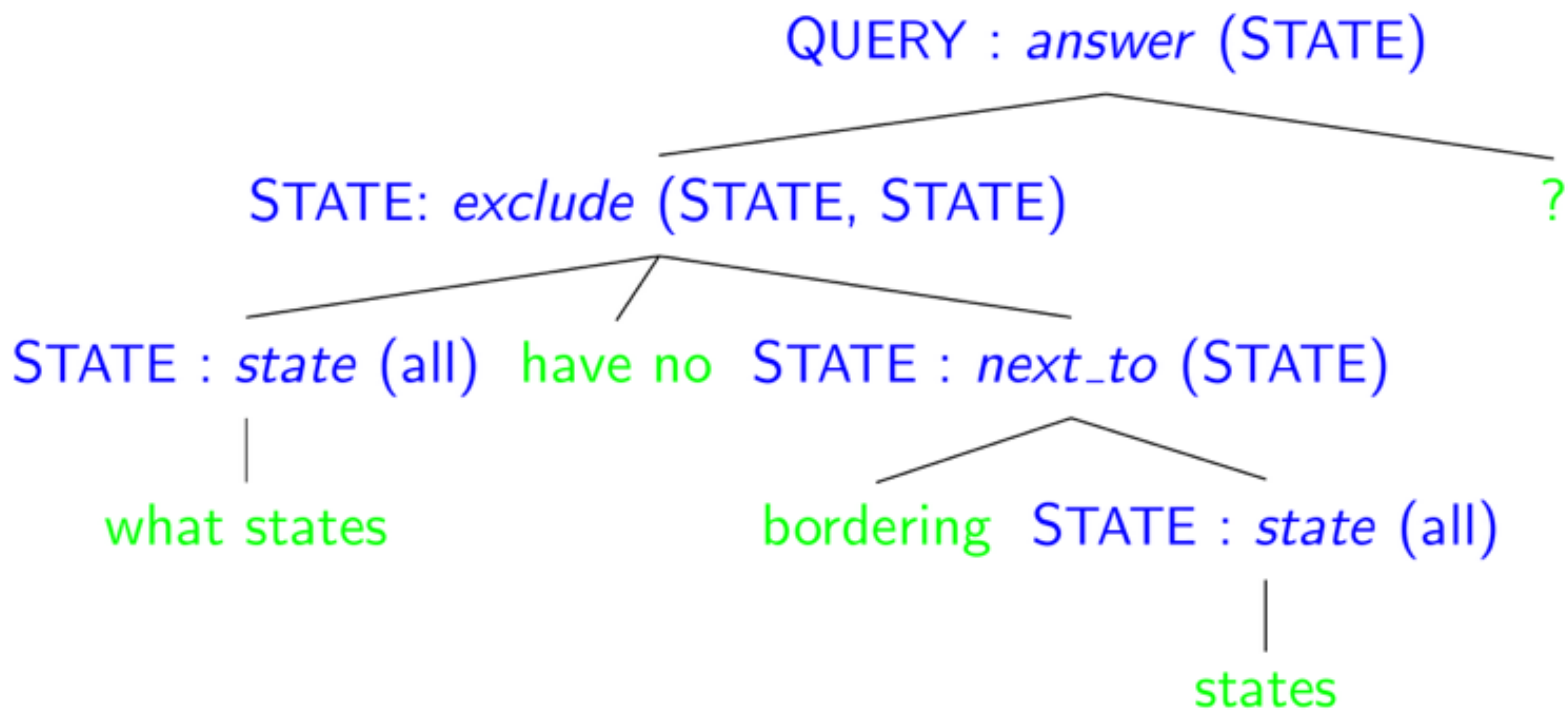
# Hybrid Tree

Input: what states have no bordering states?

(Lu et al.， 2008)

# Hybrid Tree

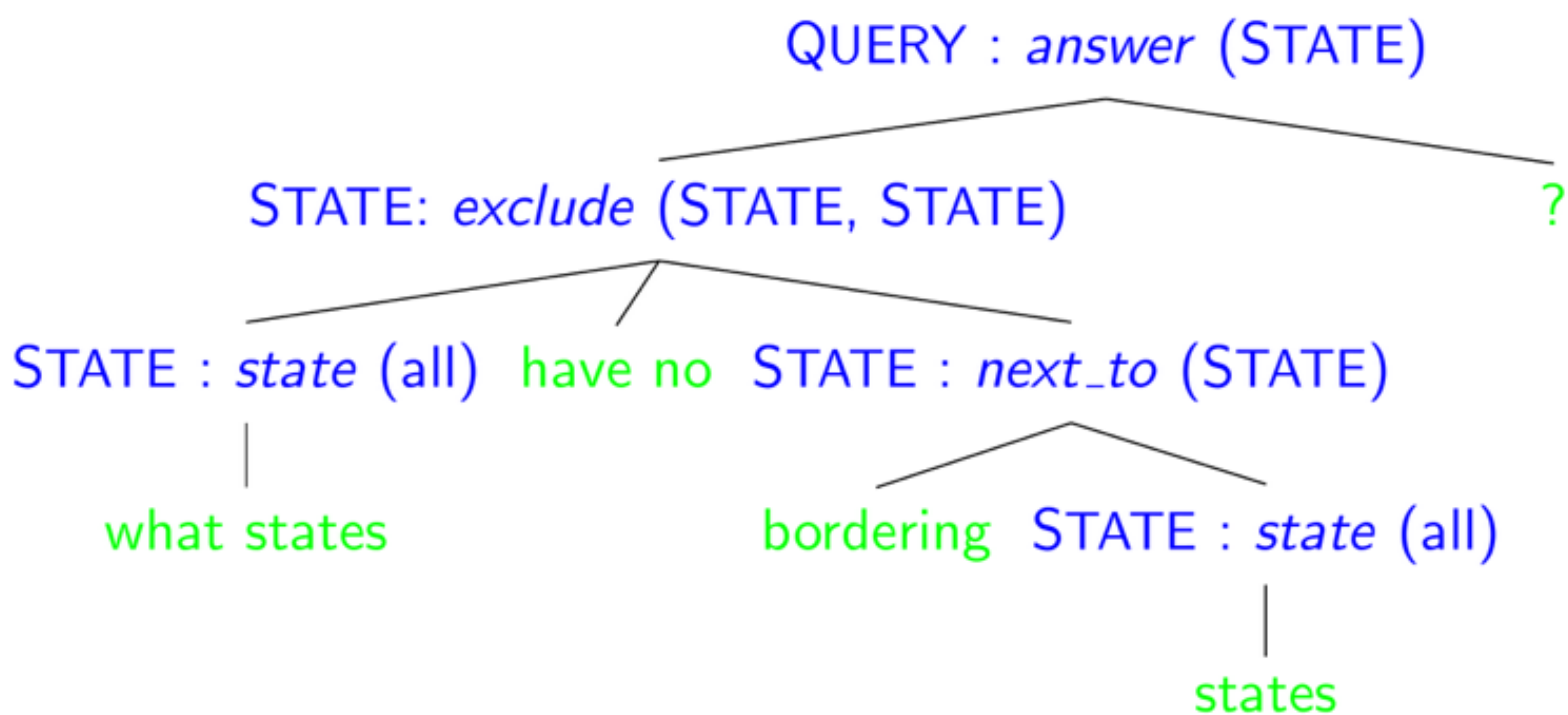
Input: what states have no bordering states?



(Lu et al., 2008)

# Hybrid Tree

Input: what states have no bordering states?

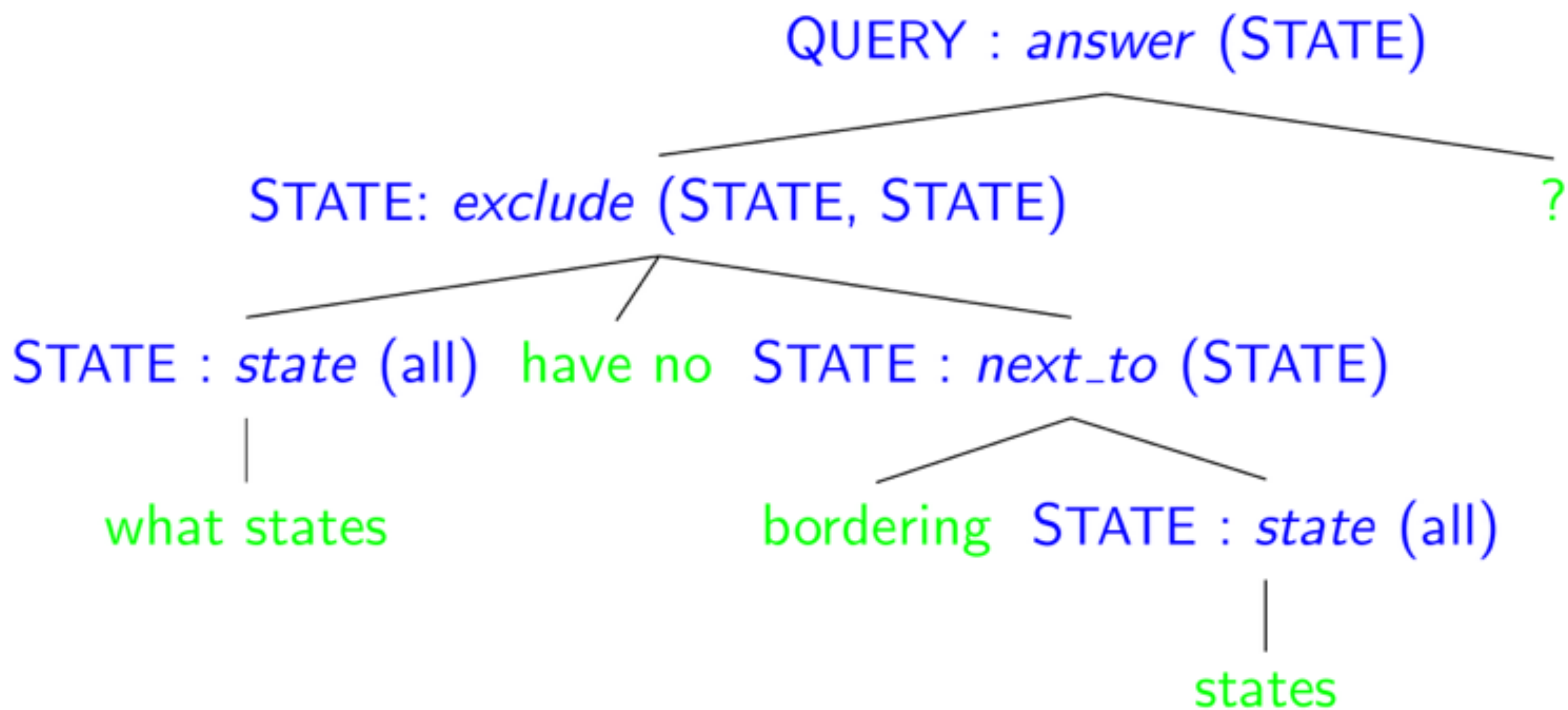



Output: $answer(exclude(state(all), next\_to(state(all))))$

(Lu et al.， 2008)

# Generative Hybrid Tree

Input: what states have no bordering states?

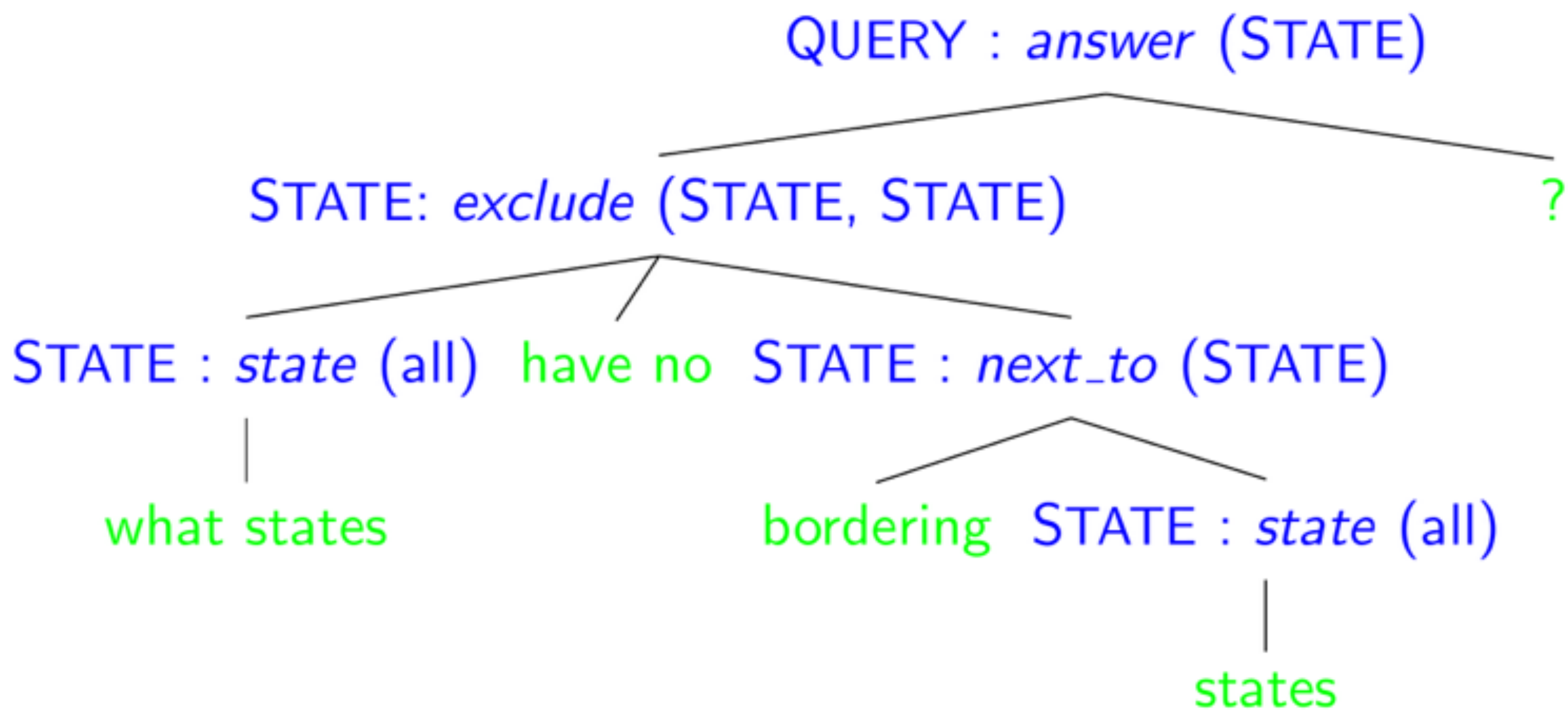



$$p(\mathbf{m}, \mathbf{n}) = \sum_{\mathbf{h} \in \mathcal{H}(\mathbf{n}, \mathbf{m})} p(\mathbf{m}, \mathbf{h}, \mathbf{n})$$

(Lu，2014, 2015)

# Discriminative Hybrid Tree

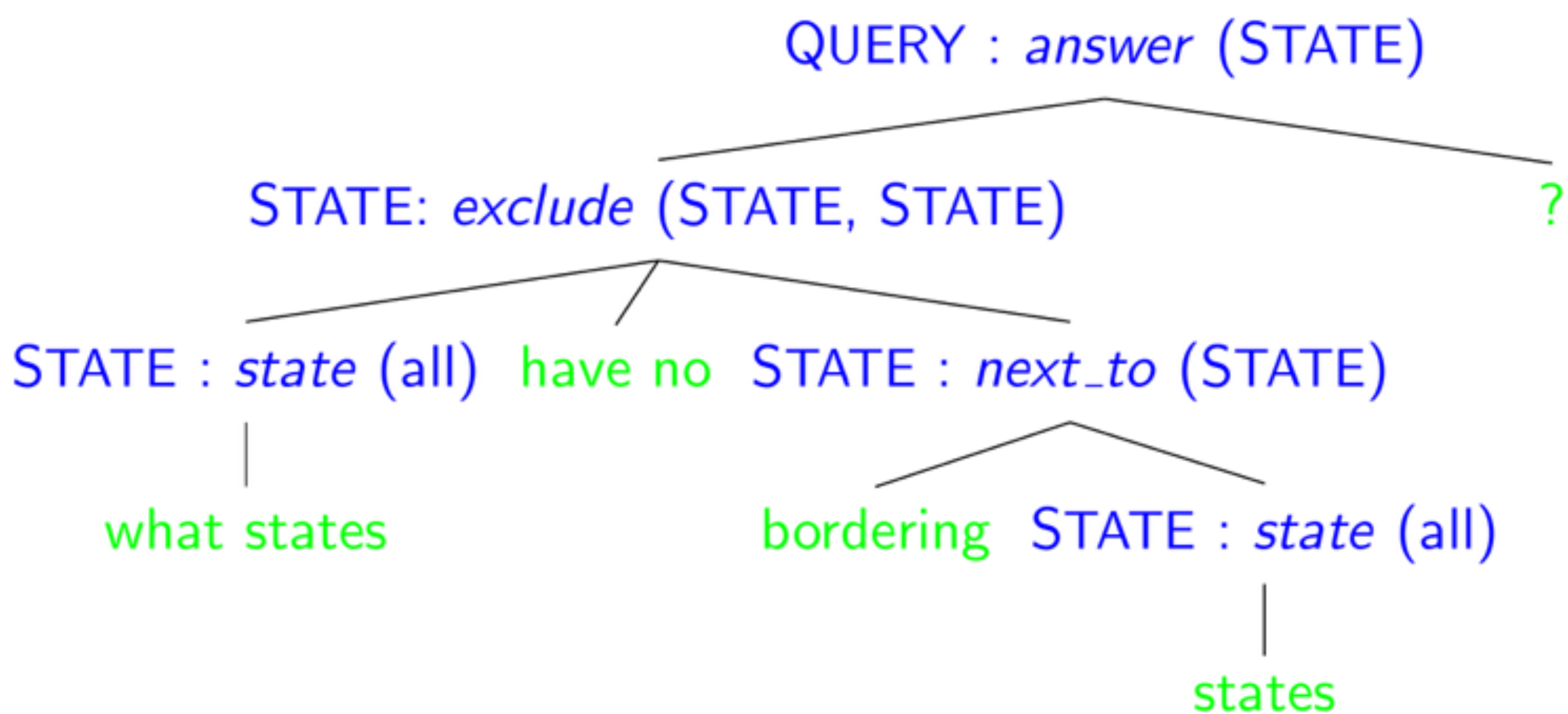
Input: what states have no bordering states?




$$p(\mathbf{m}|\mathbf{n}) = \sum_{\mathbf{h} \in \mathcal{H}(\mathbf{n},\mathbf{m})} p(\mathbf{m}, \mathbf{h}|\mathbf{n})$$

# Neural Hybrid Tree

Input: what states have no bordering states?

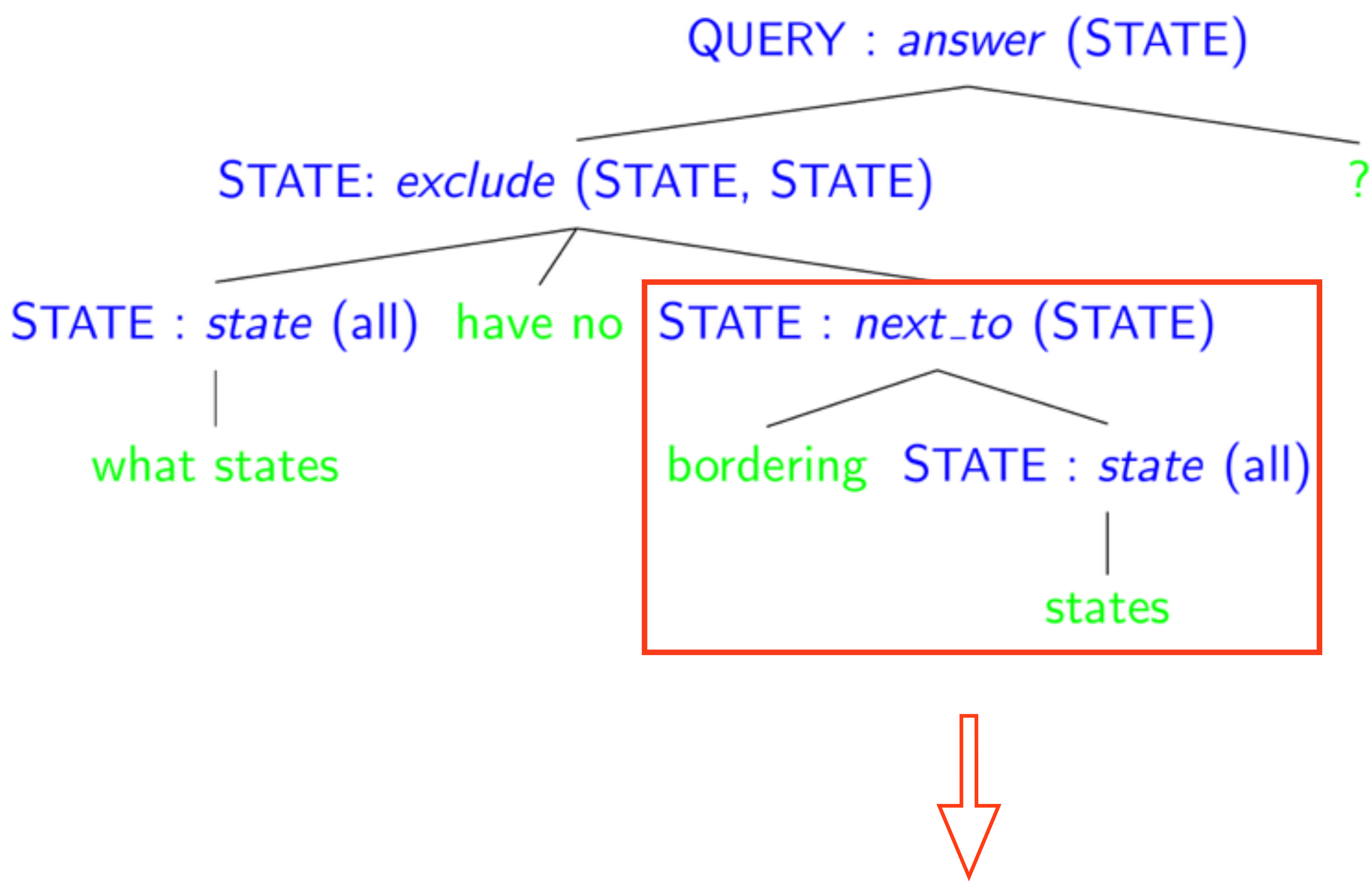



- Neural hybrid tree is an extension of discriminative hybrid tree.

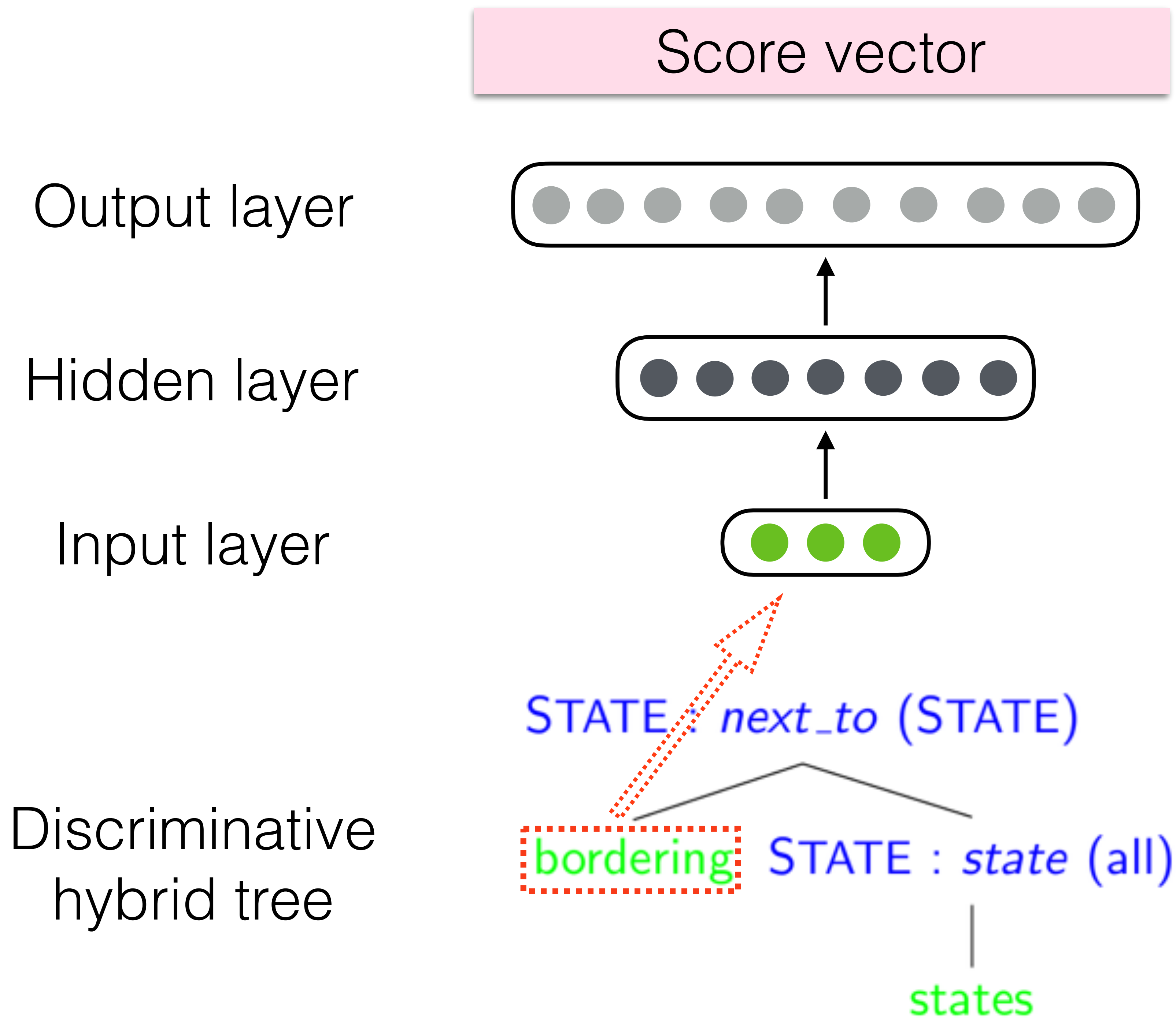
(Susanto and Lu, 2017)

# Neural Hybrid Tree

Input: what states have no bordering states?

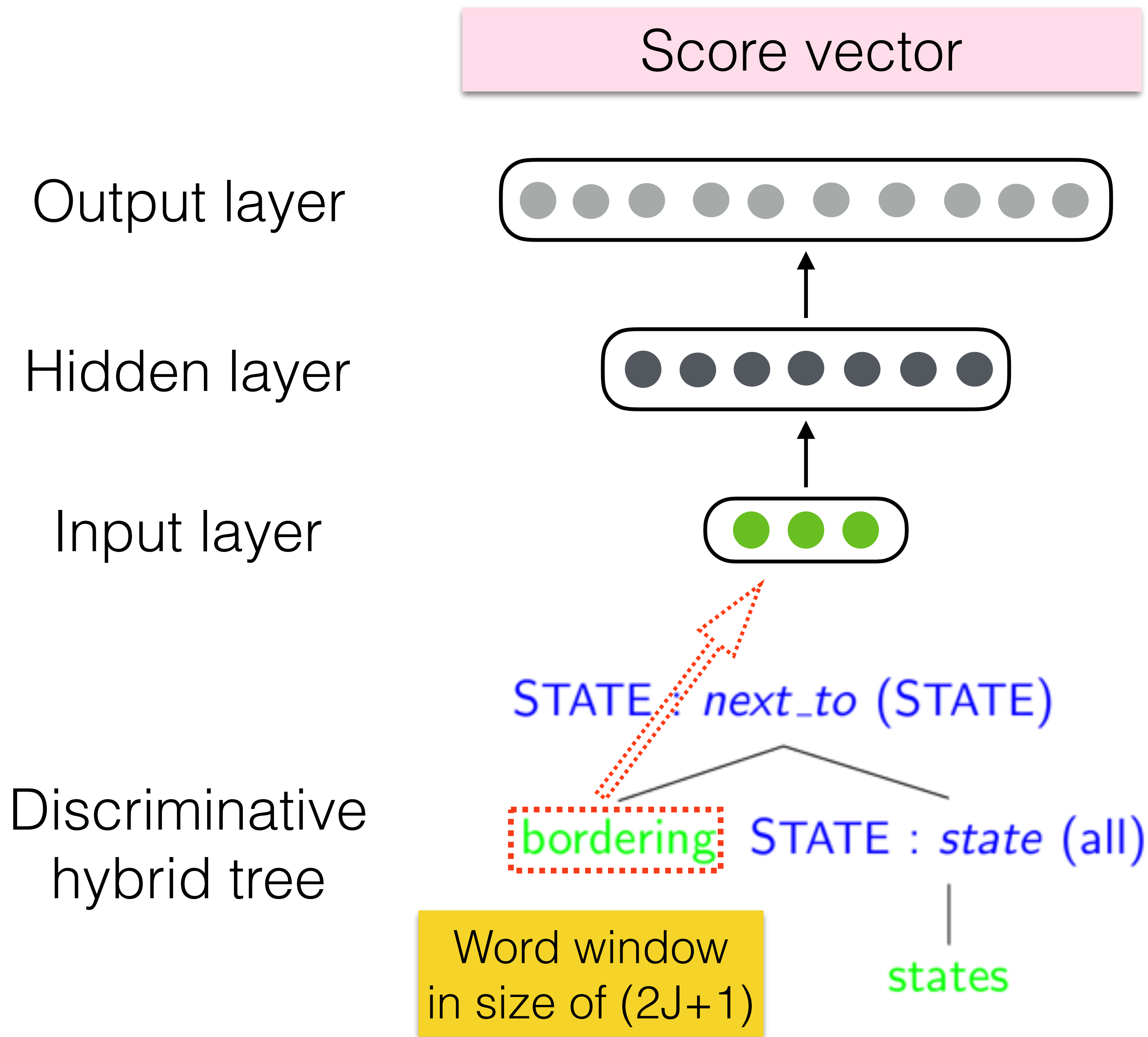


# Neural Hybrid Tree



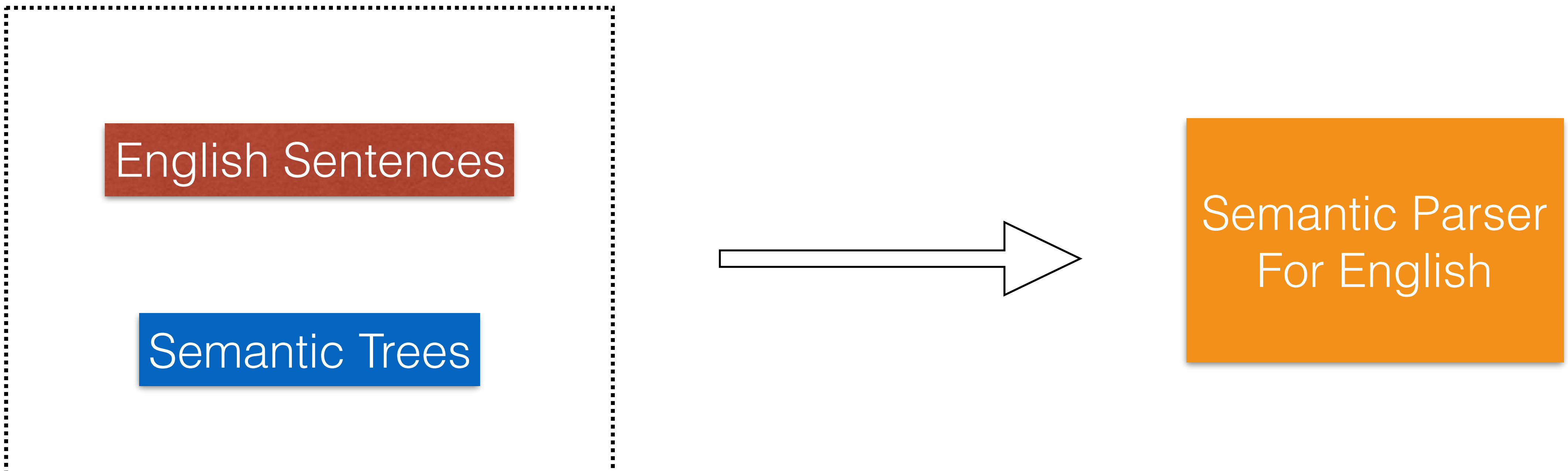
(Susanto and Lu, 2017)

# Neural Hybrid Tree



SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# What do we have?



SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# What do we have?

English Sentences

Semantic Trees

Semantic Parser For English

German Sentences

Indonesian Sentences

Chinese Sentences

⋮

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# What do we have?

English Sentences

Semantic Trees

German Sentences

Indonesian Sentences

Chinese Sentences

⋮

Semantic Parser For English

Can we leverage multi-lingual resources to improve the performance of a monolingual semantic parser?

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# What do we have?

English Sentences

Semantic Trees

German Sentences

Indonesian Sentences

Chinese Sentences

⋮

Semantic Parser For English

Can we leverage multi-lingual resources to improve the performance of a monolingual semantic parser?

The answer is Yes!!!

# Setup

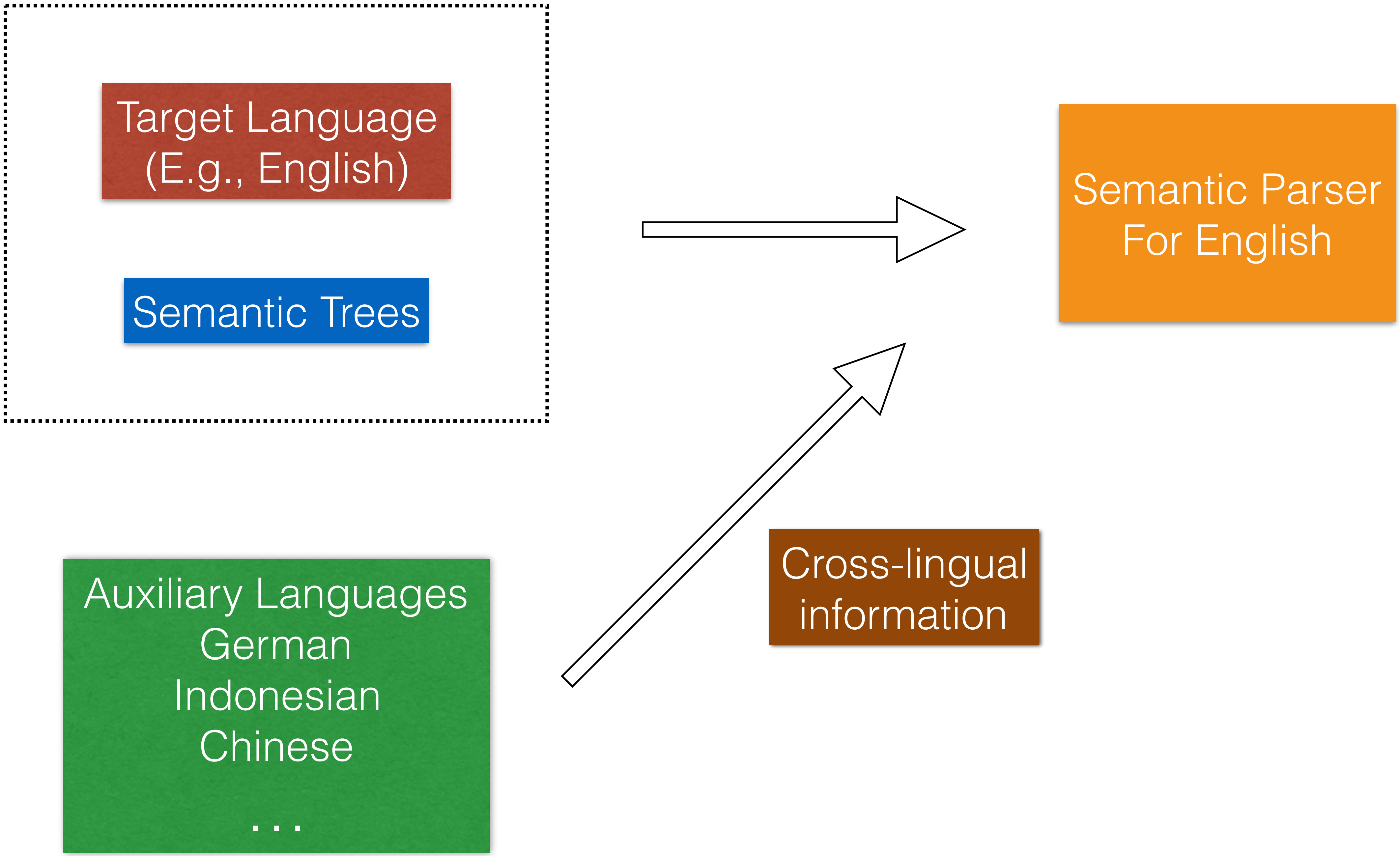
Target Language
(E.g., English)

Semantic Trees

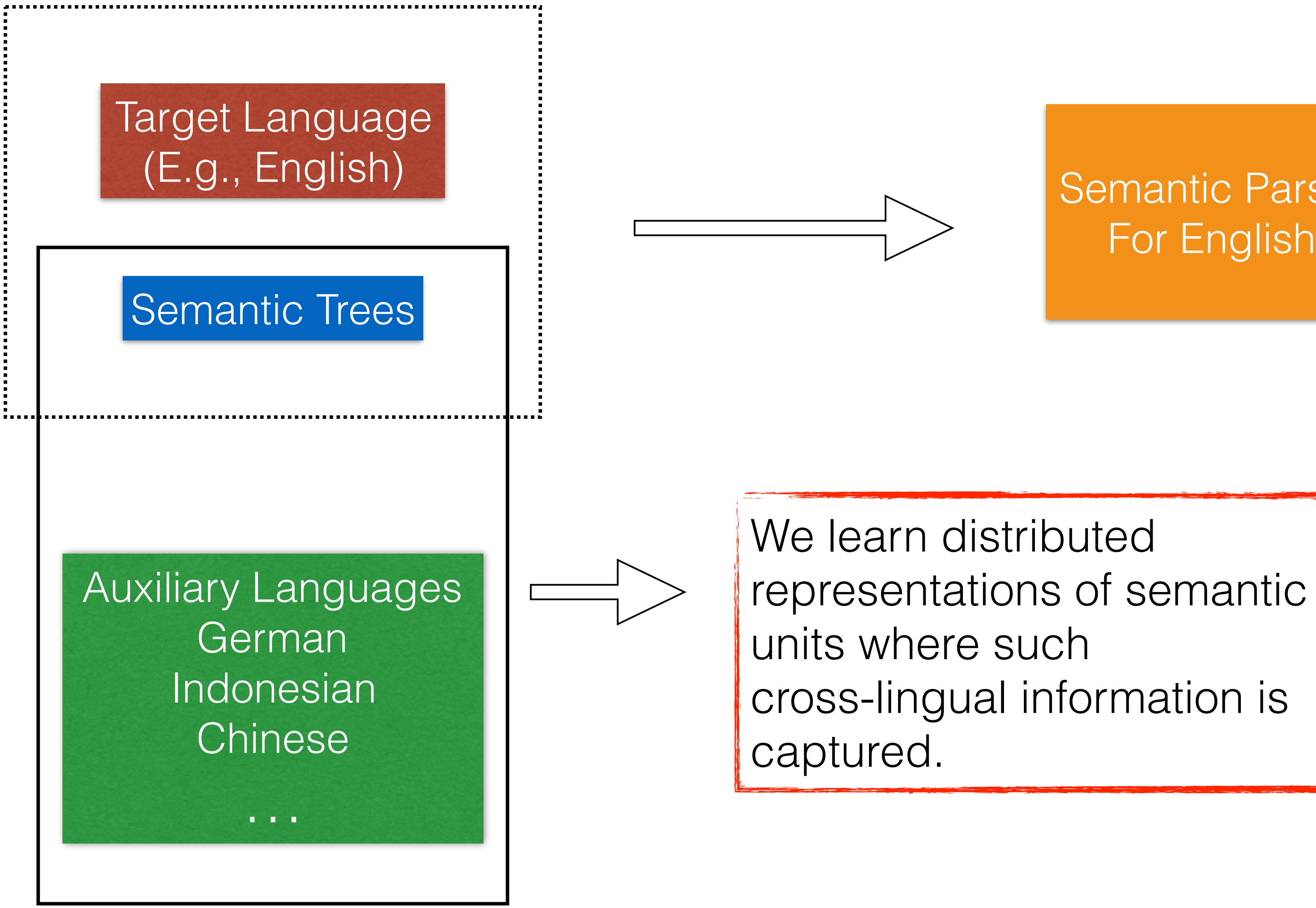
Semantic Parser
For English

Auxiliary Languages
German
Indonesian
Chinese
...

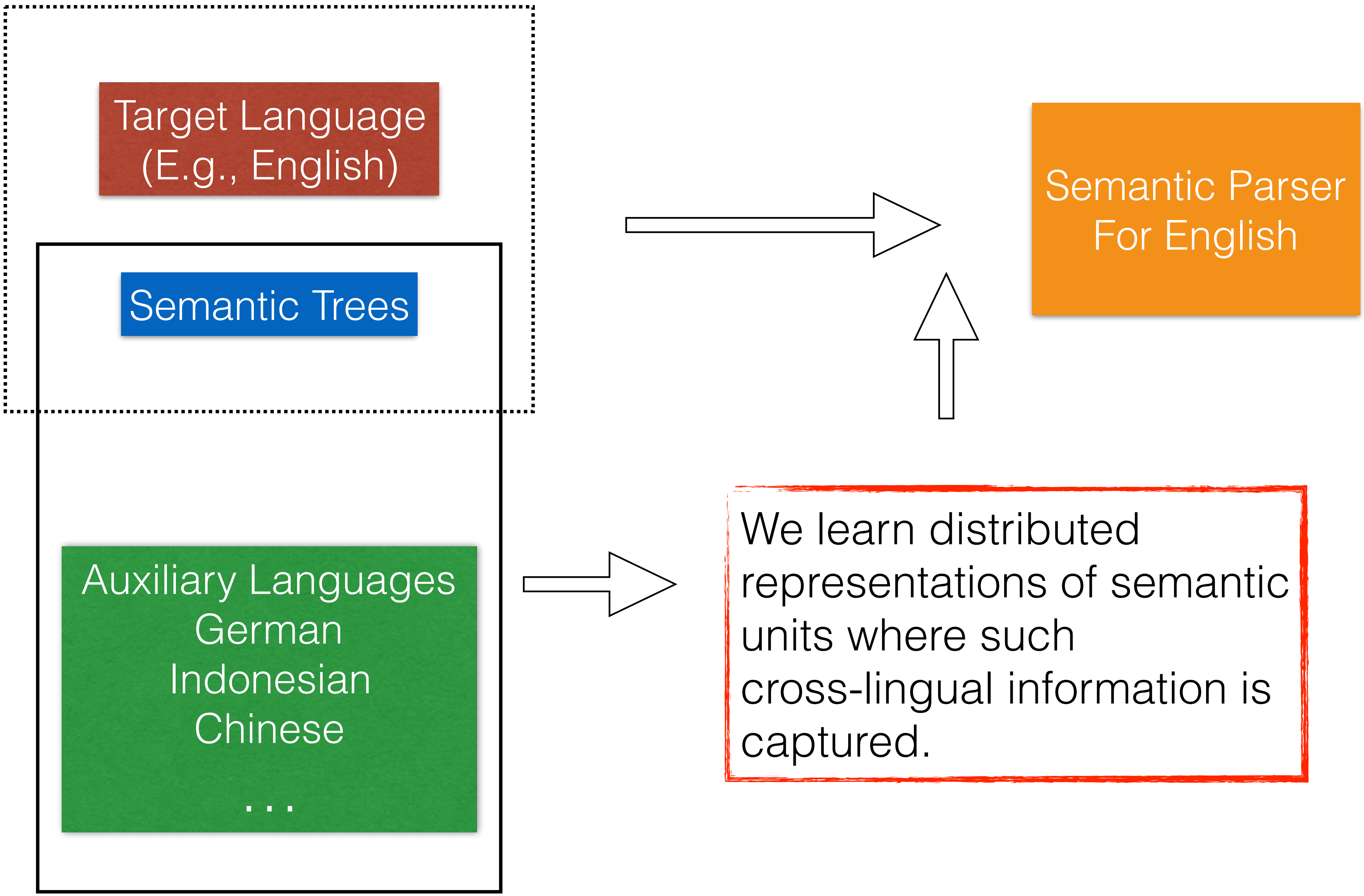
# Setup


Target Language
(E.g., English)
Semantic Trees
Semantic Parser
For English
Cross-lingual
information
Auxiliary Languages
German
Indonesian
Chinese
. . .

# Setups

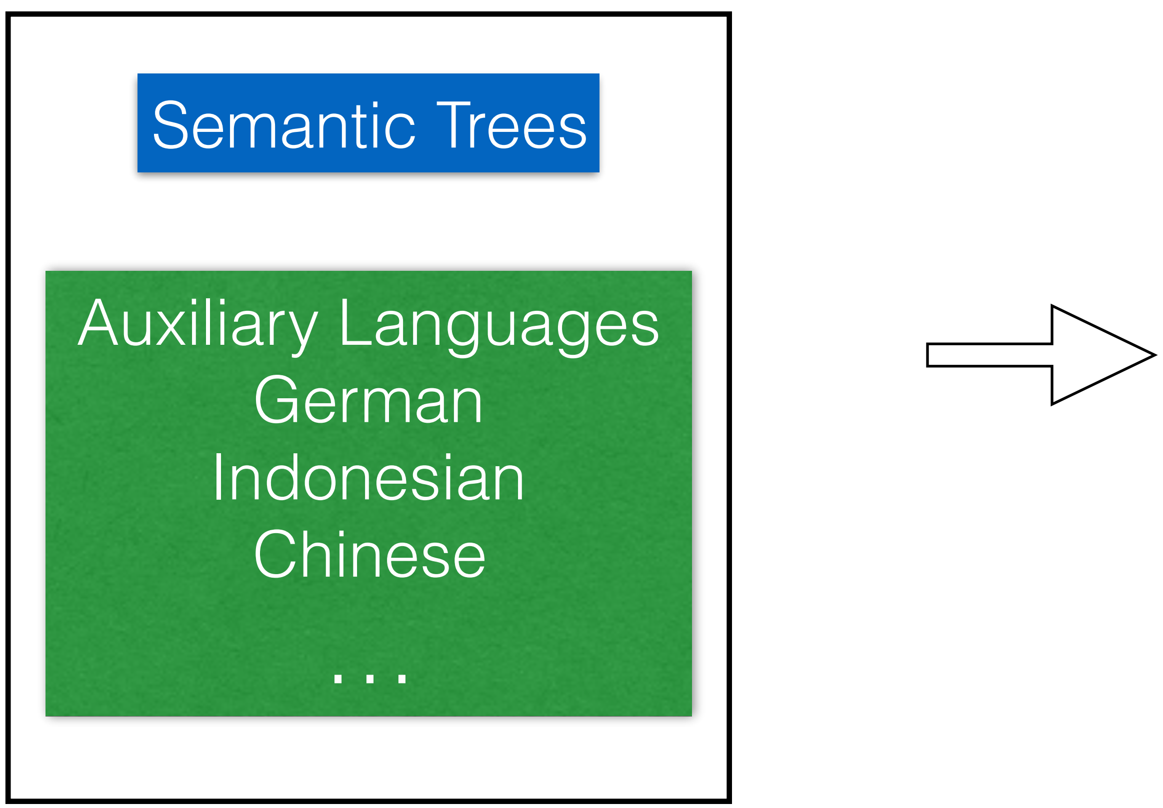

Target Language
(E.g., English)
Semantic Trees
Auxiliary Languages
German
Indonesian
Chinese
...
Semantic Parser
For English
We learn distributed representations of semantic units where such cross-lingual information is captured.

# Setups


Target Language
(E.g., English)
Semantic Trees
Auxiliary Languages
German
Indonesian
Chinese
...
Semantic Parser
For English
We learn distributed representations of semantic units where such cross-lingual information is captured.

# Cross-lingual Representations

We construct a semantics-word co-occurrence matrix $C \in R^{m \times n}$ based on auxiliary languages and semantic trees.




$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix}$$

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# Cross-lingual Representations

The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

$$C = U\Sigma V^*$$

We truncate the diagonal matrix $\Sigma$ and left multiply it with $U$ :

$$R = U\tilde{\Sigma}$$

# Cross-lingual Representations

The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

$$C = U \Sigma V^*$$

We truncate the diagonal matrix $\Sigma$ and left multiply it with $U$ :

$$R = U\tilde{\Sigma}$$

Each row in ***R*** is a d-dimensional vector, giving a low-dimensional representation for one semantic unit.

# Cross-lingual Representations

The singular value decomposition (SVD) is then applied to the co-occurrence matrix, leading to

$$C = U\Sigma V^*$$

We truncate the diagonal matrix $\Sigma$ and left multiply it with $U$ :

$$R = U\tilde{\Sigma}$$

Each row in ***R*** is a d-dimensional vector, giving a low-dimensional representation for one semantic unit.

The learned representations are considered as features for discriminative and neural hybrid tree models.

# Results

Data: Multilingual Geoquery

# Results without Neural Features

Data: Multilingual Geoquery

Baselines: (Lu et al., 2008) (Lu, 2015)

| | English | | Thai | | German | | Greek | | Chinese | | Indonesian | | Swedish | | Farsi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. |
| HT-G | 76.8 | 81.0 | 73.6 | 76.7 | 62.1 | 68.5 | 69.3 | 74.6 | 56.1 | 58.4 | 66.4 | 72.8 | 61.4 | 70.5 | 51.8 | 58.6 |
| HT-D | 86.8 | 86.8 | 80.7 | 80.7 | 75.7 | 75.7 | 79.3 | 79.3 | 76.1 | 76.1 | 75.0 | 75.0 | 79.3 | 79.3 | 73.9 | 73.9 |

# Results without Neural Features

Data: Multilingual Geoquery

Baselines: (Lu et al., 2008) (Lu, 2015)

(+o): models with distributed representations of semantic units.

| | English | | Thai | | German | | Greek | | Chinese | | Indonesian | | Swedish | | Farsi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. |
| HT-G | 76.8 | 81.0 | 73.6 | 76.7 | 62.1 | 68.5 | 69.3 | 74.6 | 56.1 | 58.4 | 66.4 | 72.8 | 61.4 | 70.5 | 51.8 | 58.6 |
| HT-D | **86.8** | **86.8** | 80.7 | 80.7 | **75.7** | **75.7** | 79.3 | 79.3 | 76.1 | 76.1 | 75.0 | 75.0 | **79.3** | **79.3** | 73.9 | 73.9 |
| HT-D (+O) | 86.1 | 86.1 | **81.1** | **81.1** | 73.6 | 73.6 | **81.4** | **81.4** | **77.9** | **77.9** | **79.6** | **79.6** | **79.3** | **79.3** | **75.7** | **75.7** |

# Results with Neural Features

Data: Multilingual Geoquery

Baselines: (Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)

(+o): models with distributed representations of semantic units.

| | **English** | | **Thai** | | **German** | | **Greek** | | **Chinese** | | **Indonesian** | | **Swedish** | | **Farsi** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. |
| HT-G | 76.8 | 81.0 | 73.6 | 76.7 | 62.1 | 68.5 | 69.3 | 74.6 | 56.1 | 58.4 | 66.4 | 72.8 | 61.4 | 70.5 | 51.8 | 58.6 |
| HT-D | **86.8** | **86.8** | 80.7 | 80.7 | **75.7** | **75.7** | 79.3 | 79.3 | 76.1 | 76.1 | 75.0 | 75.0 | **79.3** | **79.3** | 73.9 | 73.9 |
| HT-D (+O) | 86.1 | 86.1 | **81.1** | **81.1** | 73.6 | 73.6 | **81.4** | **81.4** | **77.9** | **77.9** | **79.6** | **79.6** | **79.3** | **79.3** | **75.7** | **75.7** |
| HT-D (NN) J=0 | 87.9 | 87.9 | 82.1 | 82.1 | 75.7 | 75.7 | 81.1 | 81.1 | 76.8 | 76.8 | 76.1 | 76.1 | 81.1 | 81.1 | 75.0 | 75.0 |
| HT-D (NN) J=1 | 88.6 | 88.6 | 84.6 | 84.6 | 76.8 | 76.8 | 79.6 | 79.6 | 75.4 | 75.4 | 78.6 | 78.6 | 82.9 | 82.9 | 76.1 | 76.1 |
| HT-D (NN) J=2 | 90.0 | 90.0 | 82.1 | 82.1 | 73.9 | 73.9 | 80.7 | 80.7 | 81.1 | 81.1 | 81.8 | 81.8 | 83.9 | 83.9 | 74.6 | 74.6 |

# Results with Neural Features

Data: Multilingual Geoquery

Baselines: (Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)

(+o): models with distributed representations of semantic units.

| | English | | Thai | | German | | Greek | | Chinese | | Indonesian | | Swedish | | Farsi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. |
| HT-G | 76.8 | 81.0 | 73.6 | 76.7 | 62.1 | 68.5 | 69.3 | 74.6 | 56.1 | 58.4 | 66.4 | 72.8 | 61.4 | 70.5 | 51.8 | 58.6 |
| HT-D | **86.8** | **86.8** | 80.7 | 80.7 | **75.7** | **75.7** | 79.3 | 79.3 | 76.1 | 76.1 | 75.0 | 75.0 | **79.3** | **79.3** | 73.9 | 73.9 |
| HT-D (+O) | 86.1 | 86.1 | **81.1** | **81.1** | 73.6 | 73.6 | **81.4** | **81.4** | **77.9** | **77.9** | **79.6** | **79.6** | **79.3** | **79.3** | **75.7** | **75.7** |
| HT-D (NN) J=0 | 87.9 | 87.9 | 82.1 | 82.1 | 75.7 | 75.7 | 81.1 | 81.1 | 76.8 | 76.8 | 76.1 | 76.1 | 81.1 | 81.1 | 75.0 | 75.0 |
| HT-D (NN) J=1 | 88.6 | 88.6 | 84.6 | 84.6 | ***76.8*** | ***76.8*** | 79.6 | 79.6 | 75.4 | 75.4 | 78.6 | 78.6 | 82.9 | 82.9 | 76.1 | 76.1 |
| HT-D (NN) J=2 | ***90.0*** | ***90.0*** | 82.1 | 82.1 | 73.9 | 73.9 | 80.7 | 80.7 | 81.1 | 81.1 | 81.8 | 81.8 | ***83.9*** | ***83.9*** | 74.6 | 74.6 |
| HT-D (NN+O) J=0 | 86.1 | 86.1 | 83.6 | 83.6 | 73.9 | 73.9 | 82.1 | 82.1 | 77.9 | 77.9 | 81.1 | 81.1 | 82.1 | 82.1 | 74.6 | 74.6 |
| HT-D (NN+O) J=1 | 86.1 | 86.1 | ***86.1*** | ***86.1*** | 72.5 | 72.5 | 80.4 | 80.4 | 81.4 | 81.4 | 82.5 | 82.5 | 82.5 | 82.5 | 75.7 | 75.7 |
| HT-D (NN+O) J=2 | 89.6 | 86.1 | 84.6 | 84.6 | 72.1 | 72.1 | ***83.2*** | ***83.2*** | ***82.1*** | ***82.1*** | ***83.9*** | ***83.9*** | 83.6 | 83.6 | ***76.8*** | ***76.8*** |

# Results with Neural Features

Data: Multilingual Geoquery

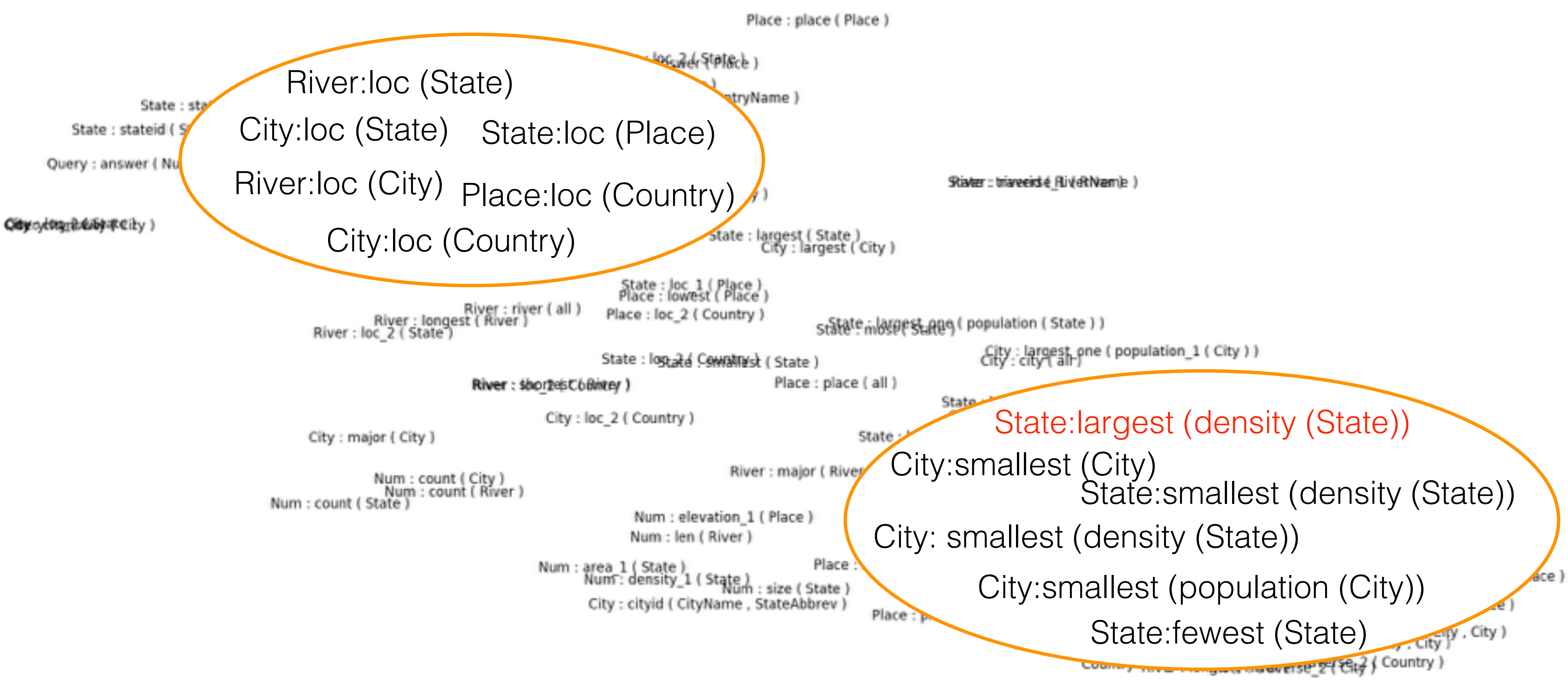
Baselines: (Lu et al., 2008) (Lu, 2015) (Susanto and Lu, 2017)

(+o): models with distributed representations of semantic units.

| | English | | Thai | | German | | Greek | | Chinese | | Indonesian | | Swedish | | Farsi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. | Acc. | F. |
| HT-G | 76.8 | 81.0 | 73.6 | 76.7 | 62.1 | 68.5 | 69.3 | 74.6 | 56.1 | 58.4 | 66.4 | 72.8 | 61.4 | 70.5 | 51.8 | 58.6 |
| HT-D | **86.8** | **86.8** | 80.7 | 80.7 | **75.7** | **75.7** | 79.3 | 79.3 | 76.1 | 76.1 | 75.0 | 75.0 | **79.3** | **79.3** | 73.9 | 73.9 |
| HT-D (+O) | 86.1 | 86.1 | **81.1** | **81.1** | 73.6 | 73.6 | **81.4** | **81.4** | **77.9** | **77.9** | **79.6** | **79.6** | **79.3** | **79.3** | **75.7** | **75.7** |
| HT-D (NN) J=0 | 87.9 | 87.9 | 82.1 | 82.1 | 75.7 | 75.7 | 81.1 | 81.1 | 76.8 | 76.8 | 76.1 | 76.1 | 81.1 | 81.1 | 75.0 | 75.0 |
| HT-D (NN) J=1 | 88.6 | 88.6 | 84.6 | 84.6 | ***76.8*** | ***76.8*** | 79.6 | 79.6 | 75.4 | 75.4 | 78.6 | 78.6 | 82.9 | 82.9 | 76.1 | 76.1 |
| HT-D (NN) J=2 | ***90.0*** | ***90.0*** | 82.1 | 82.1 | 73.9 | 73.9 | 80.7 | 80.7 | 81.1 | 81.1 | 81.8 | 81.8 | ***83.9*** | ***83.9*** | 74.6 | 74.6 |
| HT-D (NN+O) J=0 | 86.1 | 86.1 | 83.6 | 83.6 | 73.9 | 73.9 | 82.1 | 82.1 | 77.9 | 77.9 | 81.1 | 81.1 | 82.1 | 82.1 | 74.6 | 74.6 |
| HT-D (NN+O) J=1 | 86.1 | 86.1 | ***86.1*** | ***86.1*** | 72.5 | 72.5 | 80.4 | 80.4 | 81.4 | 81.4 | 82.5 | 82.5 | 82.5 | 82.5 | 75.7 | 75.7 |
| HT-D (NN+O) J=2 | 89.6 | 86.1 | 84.6 | 84.6 | 72.1 | 72.1 | ***83.2*** | ***83.2*** | ***82.1*** | ***82.1*** | ***83.9*** | ***83.9*** | 83.6 | 83.6 | ***76.8*** | ***76.8*** |

5 out of 8 languages get improved

# Cross-Lingual Representations




- Semantic units with similar meanings gather together.
- Occasionally, semantic units conveying opposite meanings are grouped together.

# Conclusions

- Summary
  - Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

# Conclusions

- Summary
  - Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.
- Future work
  - Learn representations and semantic parsers in a joint manner.

# Conclusions

- Summary
  - Presented a novel method to learning distributed representations of semantic units containing cross-lingual information.
- Future work
  - Learn representations and semantic parsers in a joint manner.
  - Investigate which languages from auxiliary corpus are the leading sources of performance gains.

Code available at: http://statnlp.org/research/sp/


**Questions?**