

# EAT2seq: controlled sentence transformation without task-specific training

## Sentence transformation

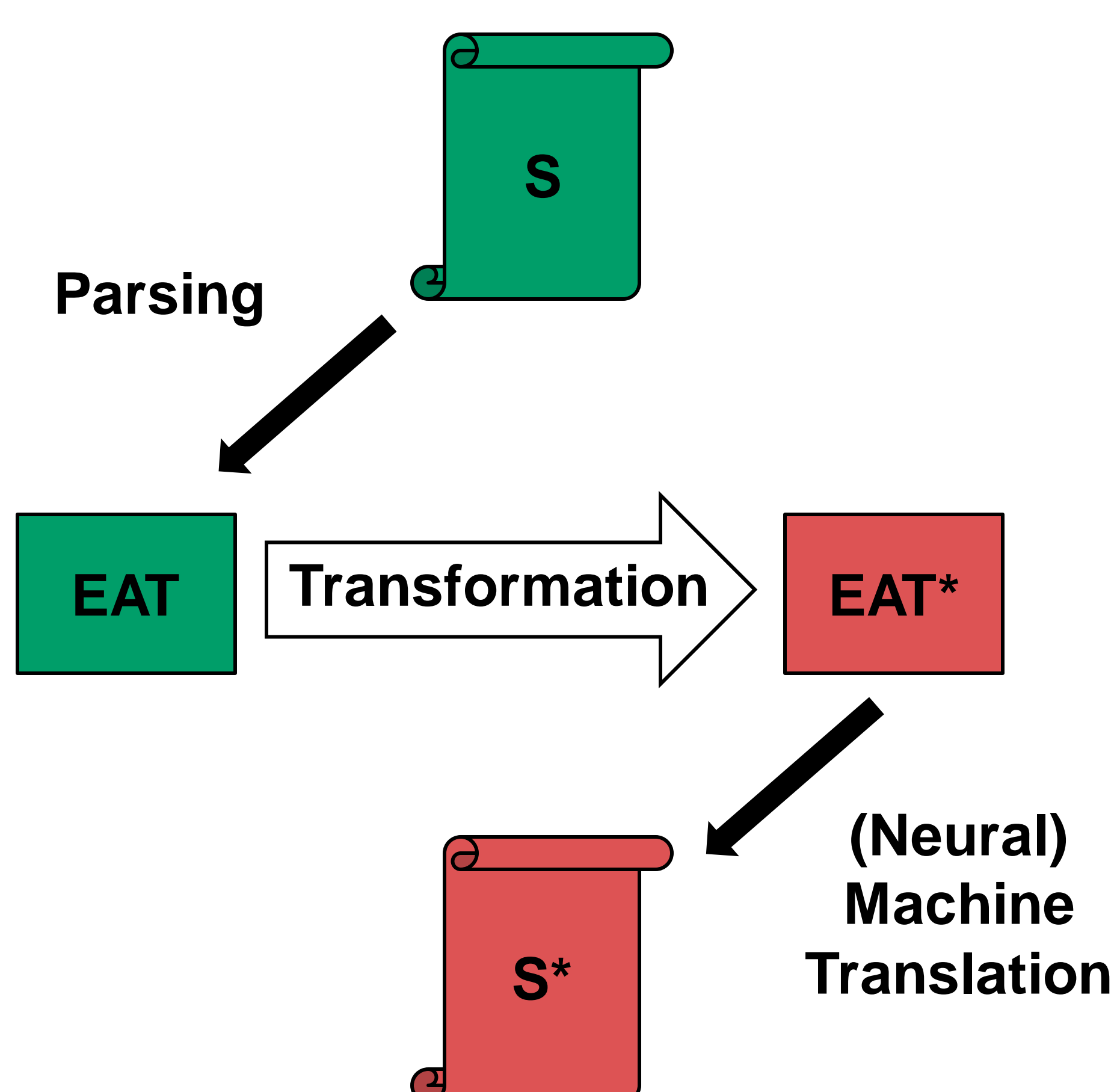
- Changing specific properties of a sentence, leaving other properties intact.
- **Parallel corpora** rarely available for training.

## EAT2seq

- **Generic** sentence transformation framework.
- **Parallel corpora not required.**
- **Controllable** by user-set parameters.
- Works in **any direction or combination.**
- Based on **EAT**: a representation of
  1. **argument structure** (Event-Agent-Theme)
  2. **grammatical information**
- Model **trained to reproduce sentence** from EAT.

## Transformation method

1. **Parse** original sentence to EAT
2. **Transform** EAT
3. **Encode** EAT with RNN (LSTM)
4. **Decode** EAT-encoding into target sentence with another RNN (LSTM)



## Experimental results

- Model trained on 8.5 million sentences.
- Transformed **300 sentences** in **14 directions**.
- **Manual evaluation** for 14 x 50 sentences
  - **66% perfect** transformations
- **Back-transformation** overlap with original
  - **48% identical** back-transformations

Category	Transformation direction	Perfect (manual)	Back-transformation	
			Identical	BLEU
Clause type	declarative-question	70%	45%	64.52
	question-declarative	74%	45%	81.82
Negation	affirmed-negated	54%	47%	67.95
	negated-affirmed	86%	55%	79.38
Voice	active-passive	48%	34%	55.29
	passive-active	44%	28%	55.55
Tense	present-past	76%	55%	75.19
	past-present	68%	49%	70.47
	present-perfect	68%	52%	72.92
	perfect-present	74%	53%	76.92
	present-pluperfect	70%	51%	73.01
	pluperfect-present	54%	41%	67.28
Aspect	perfective-imperfective	64%	50%	68.28
	imperfective-perfective	76%	52%	74.51

Evaluation of sentence transformation to **14 directions** by EAT2seq.  
**Manual evaluation** from **50 sentences** in each direction.  
**Back-transformation** evaluation from **300 sentences** in each direction.  
>50% in green, <40% in red.

## Discussion

- EAT2seq succeeded **perfectly most of the time.**
- **Poorer performance on passive transformation** likely due to **training data imbalance.**
- **Single model** used for **all transformations.**
- No manually labelled data or parallel corpora used: **training data very easy to obtain.**