# Neural sentence generation from formal semantics

Kana Manome[1], Masashi Yoshikawa[2], Hitomi Yanaka[3], Pascual Martínez-Gómez[4], Koji Mineshima[1], Daisuke Bekki[1]

[1]Ochanomizu University, [2]Nara Institute of Science and Technology, [3]RIKEN, [4]Amazon

## Abstract

- Presenting a sequence-to-sequence (seq2seq) [Sutskever et al., 2014] model with attention [Bahdanau et al., 2015] for generating sentences from logical formula based on event semantics
- Augmenting the seq2seq model with masking to constrain vocabularies of output sentences
- Using a semantic parsing system [Martínez-Gómez et al., 2016] based on Combinatory Categorial Grammar (CCG) [Steedman, 2000] to obtain pairs of sentence and logical formula
- Proposing a novel evaluation method for generation using Recognizing Textual Entailment (RTE)
- Our model outperformed a baseline with respect to both BLEU scores and accuracies in RTE

## Related Work

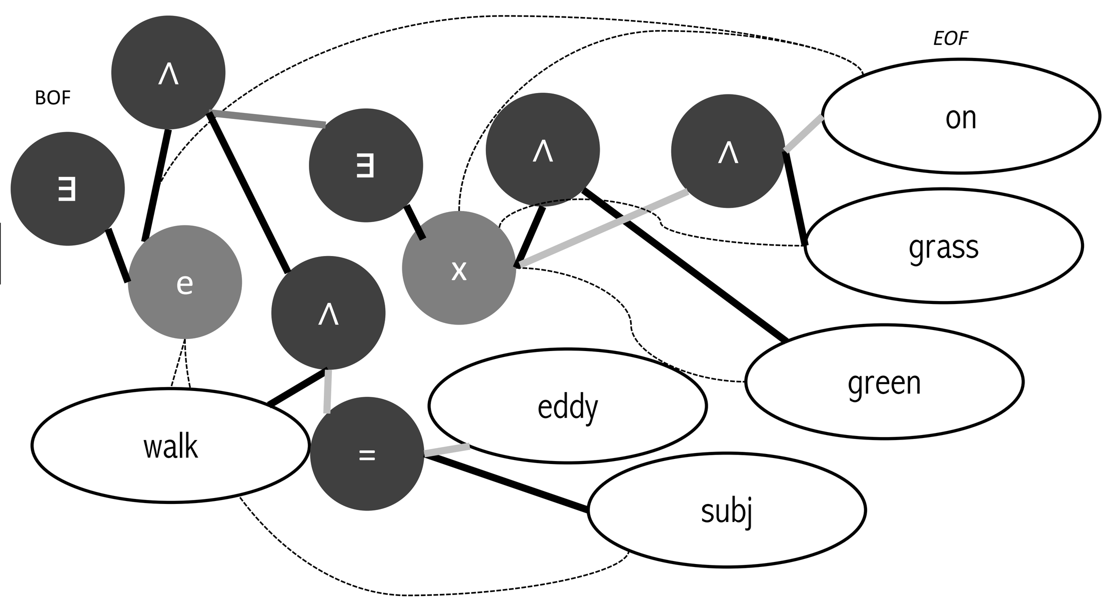
- Rule-based surface realization systems from event semantics:
  - Minimal Recursion Semantics (MRS) [Carroll and Oepen, 2005]
  - CCG [White, 2006; White and Rajkumar, 2009]
  - Treebank Semantics based on event semantics [Butler, 2016]
- ML/DL-based generation systems with semantic parsing
  - SMT-based system (GeoQuery) [Wong and Mooney, 2007]
  - Neural AMR (AMR-graph) [Konstas et al., 2017]

## Seq2seq model with attention

### Linearization

(1) Token-based linearization:
[exists, e, (, walk, (, e, ), &, Subj, (, e, ), ...]

(2) Graph-based linearization [Wang et al. ,2017] :
[exists, e, &, &, walk, =, Subj, eddy, ...]




### Masking output probabilities

1. Preparing a mapping from predicates to their realized forms

formula ∃e.( walk (e) & (Subj(e) = eddy) & ∃x.(green(x) & grass(x) & on(e, x)))

2. Verbs are mapped to their inflected forms

dict 1 {walk:[walk, walks, walked, walking}

3. Setting 1 in mask vector at positions that correspond to predicates and their inflected forms

mask 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 …
walk walks walked walking eddy

4. Functional words are always available using a predefined list of those words

dict 2 {and, both, do, every, is, no, some, the, to, will, without …}

mask 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 …
walk walks walked walking eddy and the

## Logical Formula

- Neo-Davidsonian event semantics [Parsons, 1990]

**sentence** : Eddy walked on the green grass.
**formula** : ∃e. (walk(e) & (subj(e) = eddy) & ∃x. (green(x) & grass(x) & on(e, x)))

## Experiment

### Dataset

- Creating a dataset from the SNLI corpus [Bowman et al., 2015]
- Using 50k hypothesis sentences and splitting them into train:42k/val:4k/test:4k
- Mapping the sentences into logical formulas using ccg2lambda [Martínez-Gómez et al., 2016]
- Using C&C parser for converting sentences into CCG trees

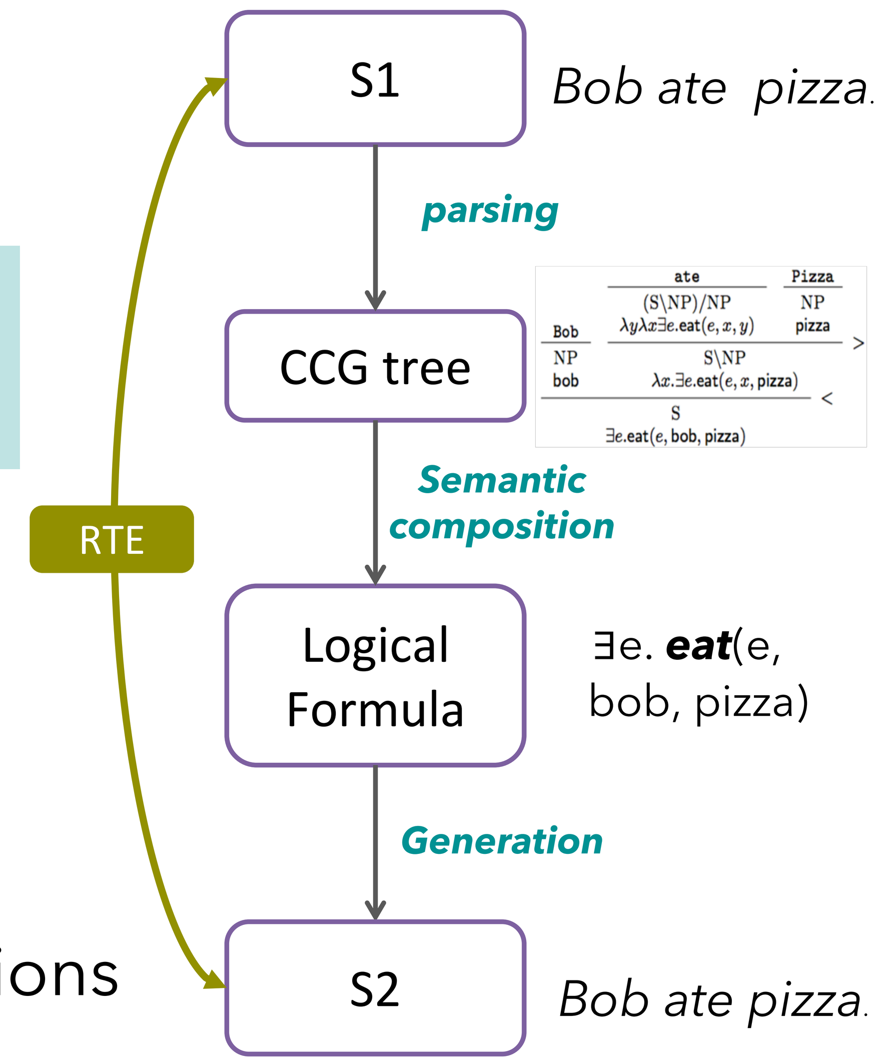
### Evaluation

- BLEU does not evaluate meaning preservation

BLEU score is high but no entailment!

***S1: No one visited the old man to greet him.***
no entailment | no entailment
***S2: Someone visited the old man to greet him.***

- We propose an evaluation method using parsing and RTE
- Parsing an input sentence S1 to obtain a formula P and then generating a sentence S2 from the formula P
- Checking whether S1 entails S2 and vice versa: S1 ⇒ S2, S2 ⇒ S1 and S1 ⇔ S2
- Using ccg2lambda for parsing and proving entailment relations




### Result

- The baseline **rule** denotes the Treebank Semantics system [Butler, 2016]
- For the RTE accuracy, the increase in the score of the **graph + mask** model was slightly larger than the increase for the **token + mask** model

| | BLEU | $S_1 \Rightarrow S_2$ | $S_2 \Rightarrow S_1$ | $S_1 \Leftrightarrow S_2$ |
|---|---|---|---|---|
| **token** | 43.0 | 87.3 | 87.3 | 87.3 |
| **+mask** | **60.0** | 92.3 | 90.8 | 89.8 |
| **graph** | 42.2 | 86.3 | 90.0 | 86.3 |
| **+mask** | 50.0 | **92.5** | **92.3** | **90.8** |
| **rule** | 38.3 | 61.5 | 62.3 | 58.8 |

| | Input sentence ($S_1$) | Decoded sentence ($S_2$) |
|---|---|---|
| (1) | the girls are swimming in the ocean. | the girls are swimming in the ocean. |
| (2) | a dog is playing fetch with his owner. | a dog is playing fetch with owner. |
| (3) | a man is sitting on the couch. | the men are sitting on a couch. |
| (4) | a tall man. | the man is tall. |
| (5) | a child is standing. | the children are standing together. |
| (6) | there are several people in this picture. | people are pictured in a picture. |

## Future Work

- Refining our model for generation of longer sentences
- Testing formulas with richer semantic information (the definite-indefinite and singular-plural distinctions for NPs and tense/aspect for VPs)

**Acknowledgement** This work was partially supported by JST AIP-PRISM Grant Number JPMJCR18Y1, Japan.