Word Representations & Dependency Parsing

Sorami Hisamoto, Kevin Duh, Masashi Shimbo & Yuji Matsumoto
Nara Institute of Science and Technology

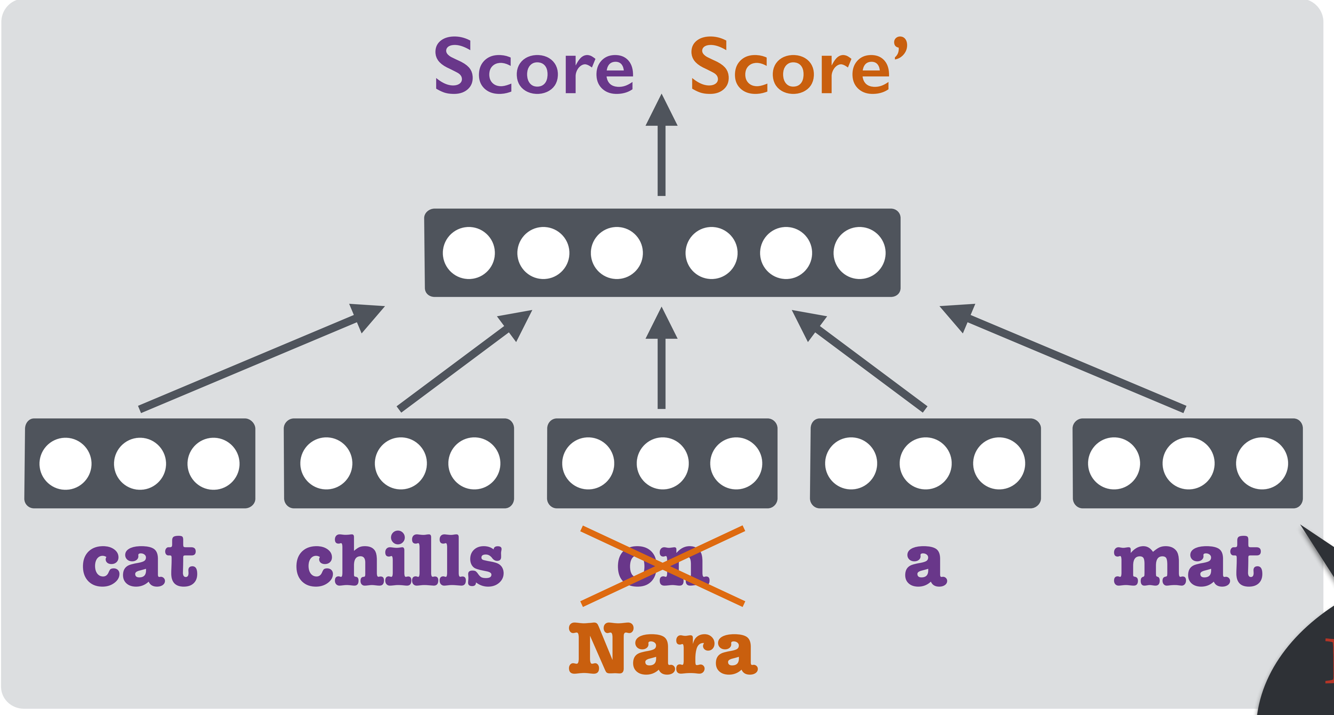
Background: Unsupervised Representation Learning

[Bengio+ 2003] [Collobert&Weston 2008] [Brown+ 1993] [Turian+ 2010] [Yu+ 2013]

Representations from large-scale unlabeled data are useful as task-independent features.

Neural Language Model

Score(Original) > Score(Corrupted) to learn weights & embeddings.

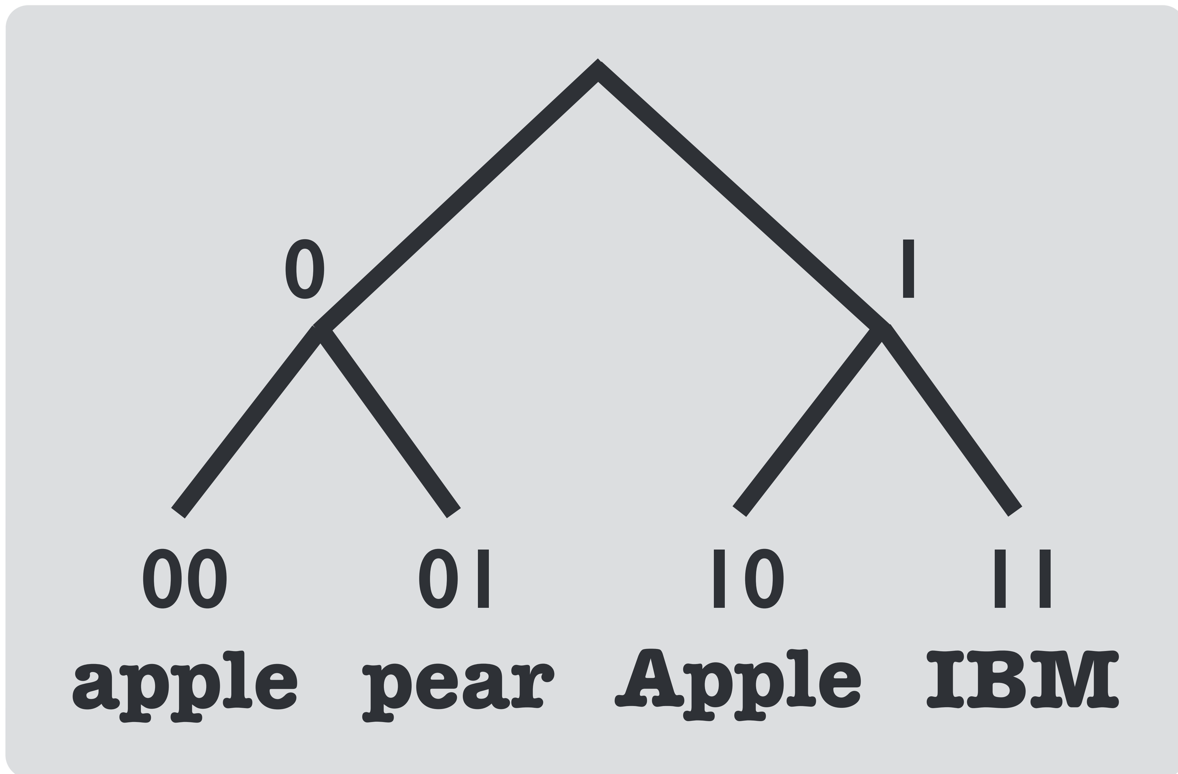


Create a corrupted sentence by replacing a word.

Words with similar contexts are close to each other

Brown Clustering

Hierarchical clustering based on class-based language model.



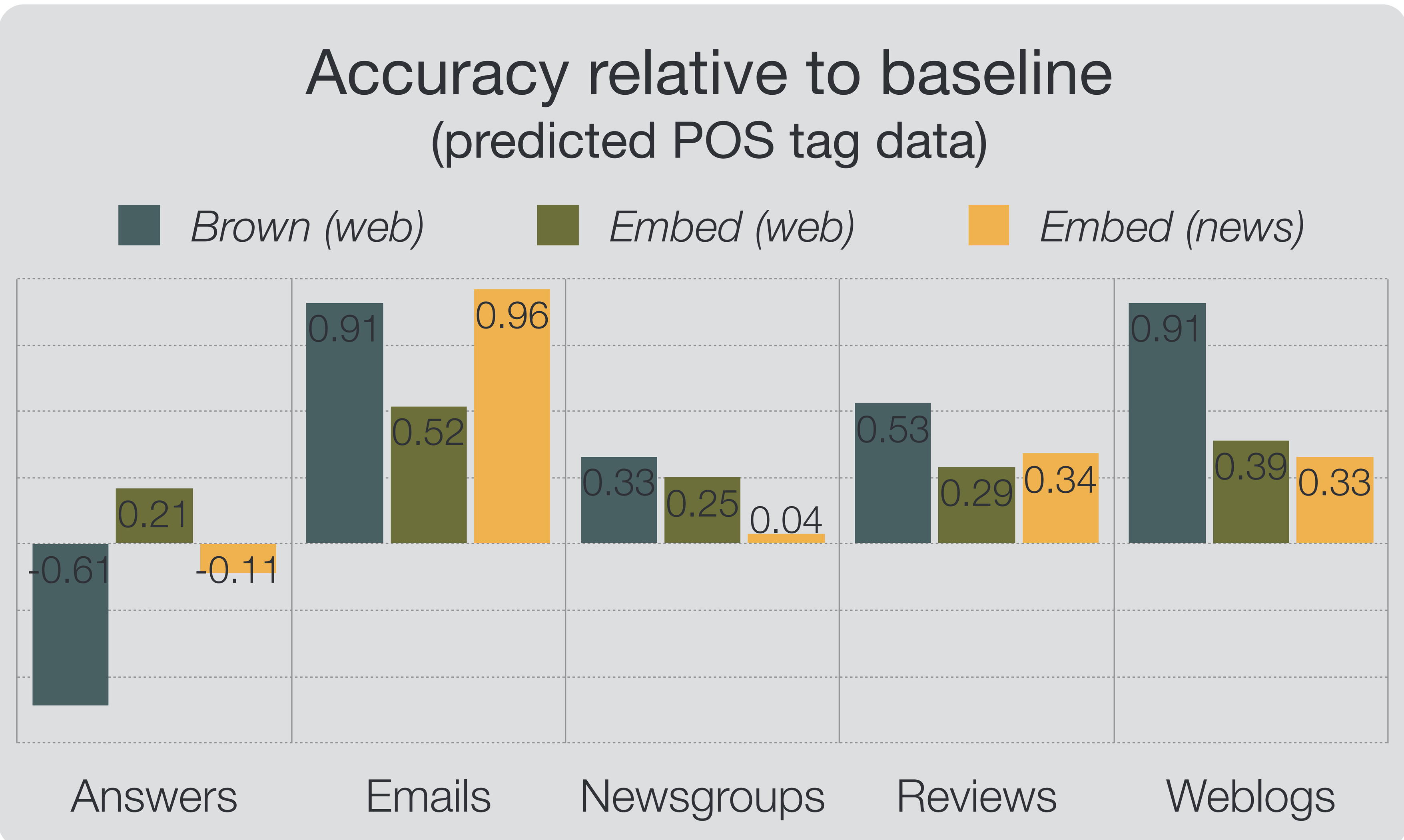
Dependency Parsing with Word Representations

Related Works [Koo+ 2008] [Wu+ 2013]

- Word reps as features for dependency parsing
- Web: small labeled data, large unlabeled data

Experiment

- Data: Google Web Treebank
- Graph-based parser (1st-order)



- Cluster embeddings to use as features

1. Embed(news): trained with newswires
2. Embed(web): trained with g-web data

Conclusion

- Results improved with predicted POS tag data, but not with gold POS.

Future Works

- Other ways to add embeddings?
- Accuracy vs. training time?

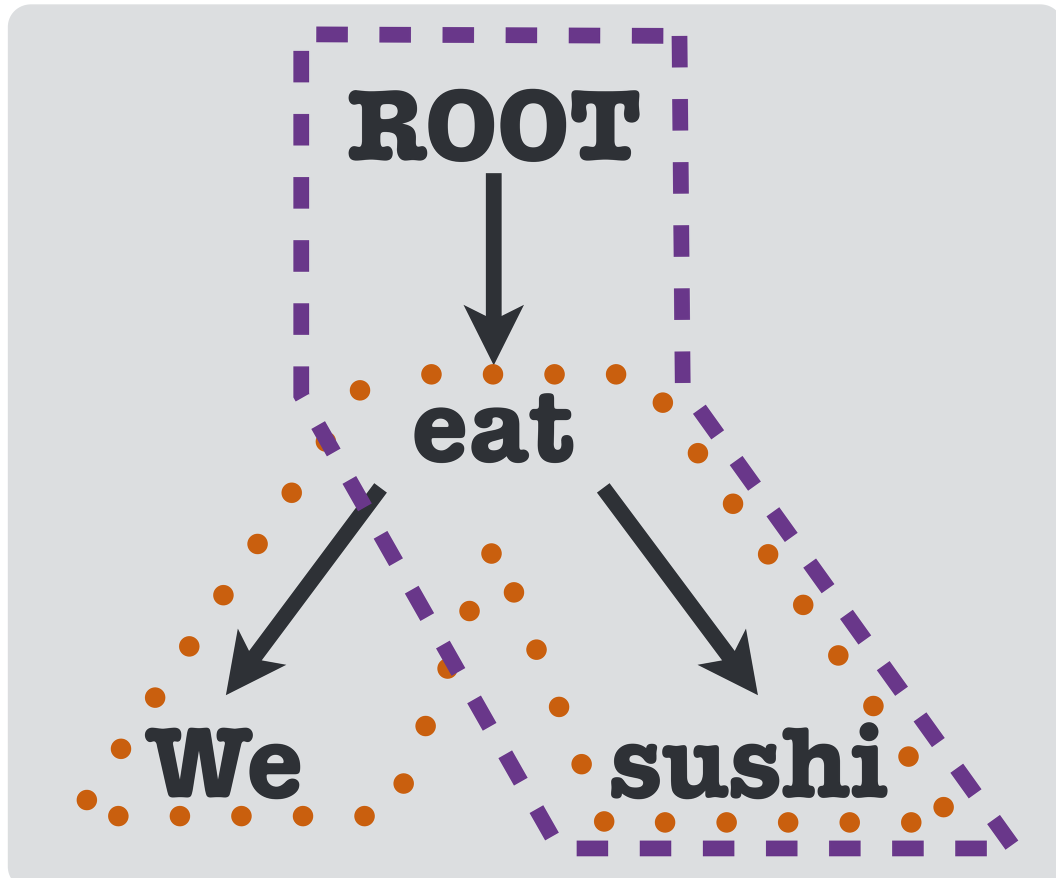
Learning Representations with Dependency Information

Related Works [Bengio+ 2003] [C&W 2008]

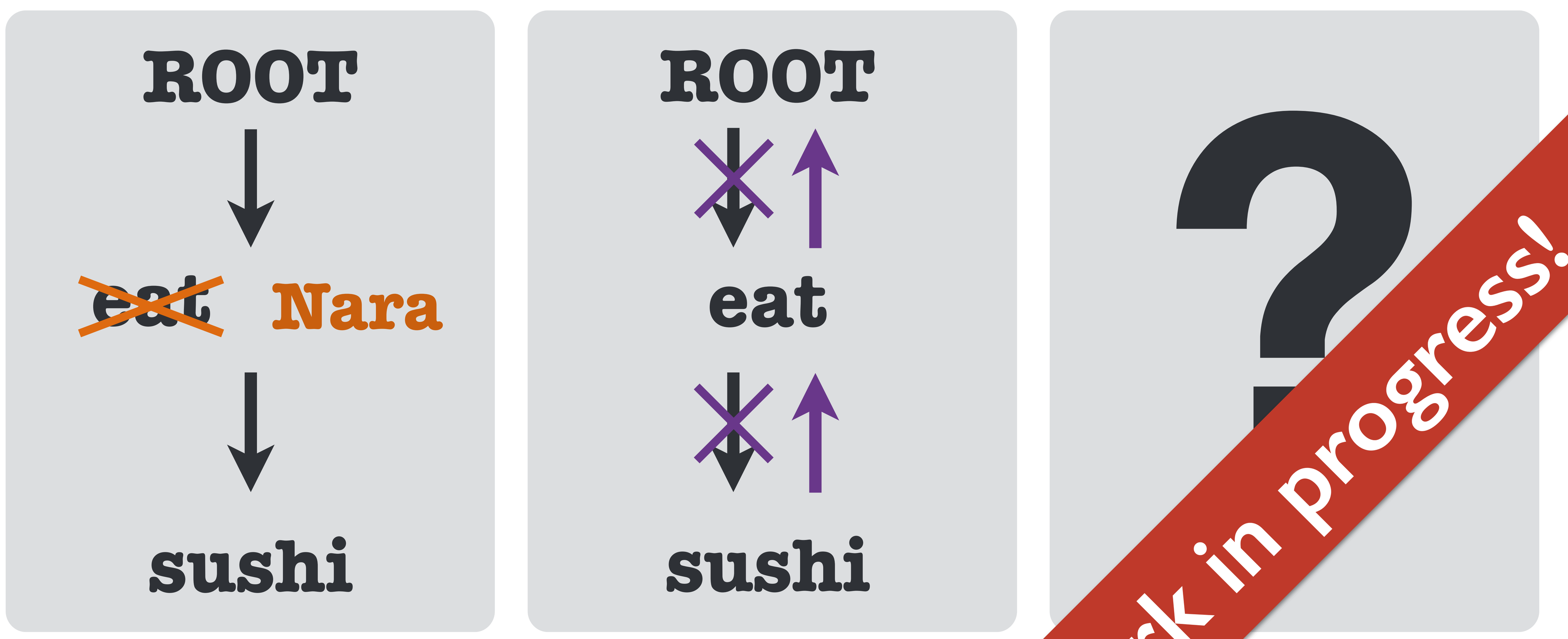
- NLM & Brown clustering use word sequences
- Can we learn better reps with dependency info?

How to input dependency information?

- Tree paths?
- e.g. trigram
  - ancestors?
  - siblings?



How to make pseudo-negative examples?



replace word

flip edge