



Learning Monolingual Compositional Representations via Bilingual Supervision



Ahmed Elgohary and Marine Carpuat, Computer Science Department, University of Maryland, College Park

Overview

Q: Can we learn to compose words into sentence representations using parallel text as supervision?

A: yes, and bilingual phrases outperform bilingual sentences and monolingual paraphrases

Background: learning sentence representations

- from PPDB paraphrases can outperform complex (un)supervised models (Paragram, Wieting et al'16)
- from bilingual sentence pairs can be useful for cross-lingual document classification (e.g. BiCVM, Hermann and Blunsom'14)

Models

- Composition by word-averaging (Wieting et al'16)

$$g(\text{piggy saw kermit dancing}) = \frac{1}{4} (\text{piggy} + \text{saw} + \text{kermit} + \text{dancing})$$

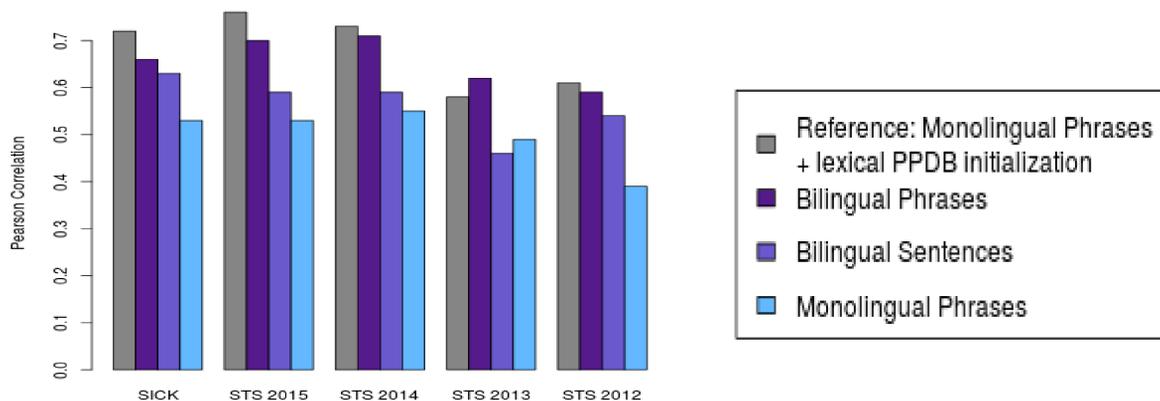
- Learn word representations using semantically equivalent text segment pairs $\langle x_1, x_2 \rangle$

$$\min_{\langle x_1, x_2 \rangle} [\delta + \underbrace{\|g(x_1) - g(x_2)\|^2}_{\text{equivalent examples closer}} - \underbrace{\|g(x_1) - g(\bar{x}_2)\|^2}_{\text{far apart from randomly selected negative examples}}]_+$$

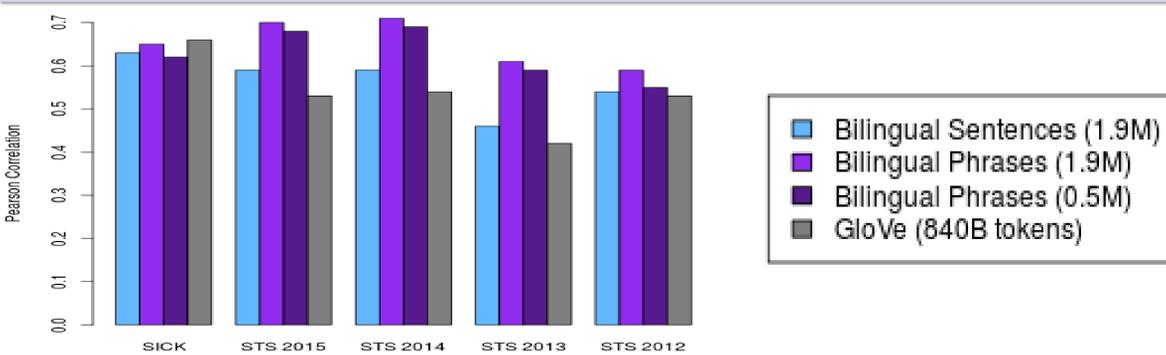
	x_1	x_2
Monolingual phrases	by our fellow member	by our colleague
Bilingual sentences	thus in fact, we might ... our fellow member	as que podramos ... nuestro colega diputado
Bilingual phrases	by our fellow member	de nuestro colega diputado

Findings

1. Bilingual phrases outperform monolingual paraphrases and parallel sentences in controlled settings



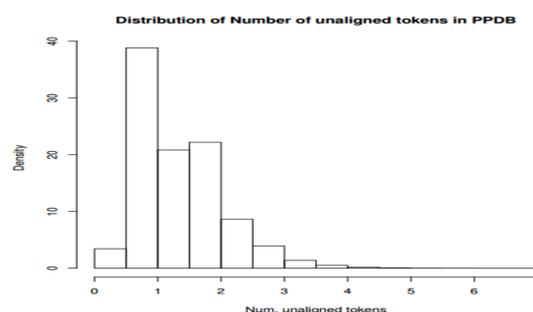
2. Training with a compositional objective requires less data



3. Bilingual phrases avoid the lexical overlap issue in paraphrases

$$\|(\text{healthy} + \text{and} + \text{stable}) - (\text{healthy} + \text{and} + \text{steady})\|$$

$$= \|\text{stable} - \text{steady}\|$$



See paper for further analysis

Evaluation Setup

2 sentence-level similarity tasks

SemEval STS (2012 - 2015)
(19 subtasks: tweets, news, webforums, image captions)

+
SICK (SemEval'14)
(designed for evaluating composition models)

Evaluation metric: Pearson correlation of cosine similarity & gold scores

3 Training Conditions	#pairs	Avg. Len	Provenance
Bilingual sentences	1.9M	28	en-es europarl-v7
Bilingual phrases	3M	5	+ Moses phrase extraction
Monolingual phrases	3M	3	XL PPDB

References

Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch "PPDB: The paraphrase database". In HLT-NAACL, 2013.

Karl Moritz Hermann and Phil Blunsom, "Multilingual models for compositional distributed semantics". In ACL, 2014.

John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu, "Towards universal paraphrastic sentence embeddings". In ICLR, 2016.