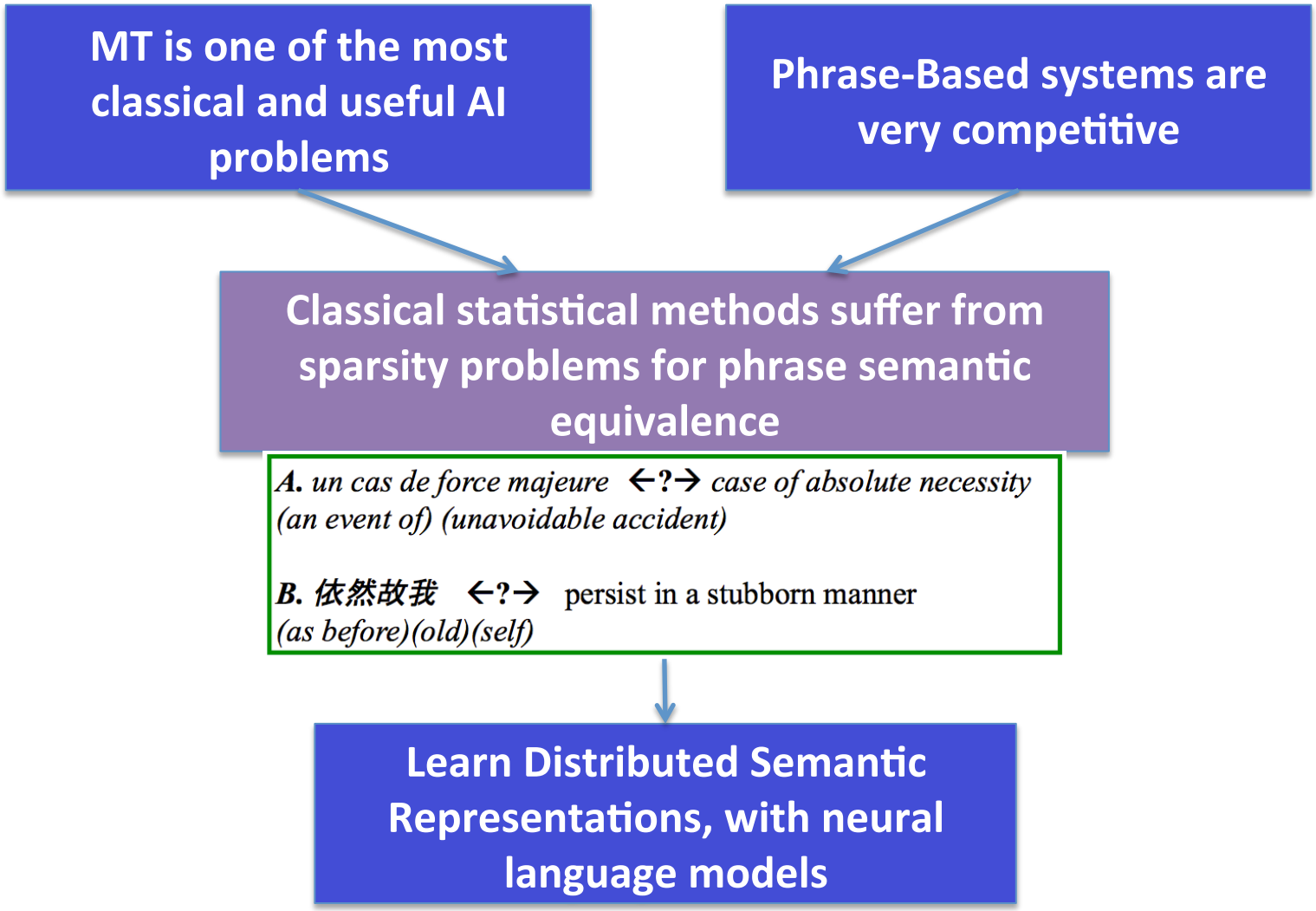
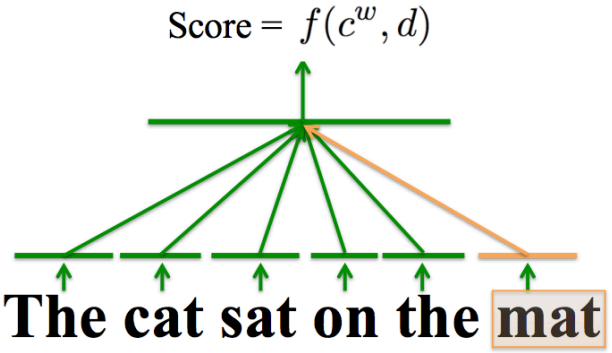
Bilingual Embeddings for Phrase-Based Machine Translation

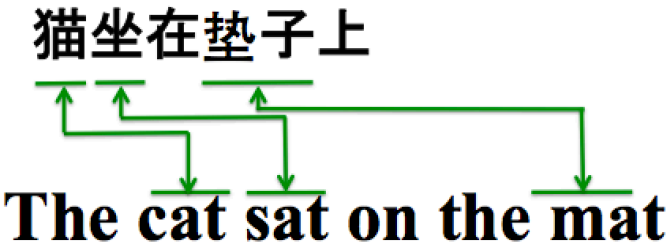
Will Y. Zou
Richard Socher
Daniel Cer
Christopher D. Manning
Stanford University

Motivation


MT is one of the most classical and useful AI problems
Phrase-Based systems are very competitive
Classical statistical methods suffer from sparsity problems for phrase semantic equivalence
A. un cas de force majeure ←?→ case of absolute necessity
(an event of) (unavoidable accident)
B. 依然故我 ←?→ persist in a stubborn manner
(as before)(old)(self)
Learn Distributed Semantic Representations, with neural language models

Model Description





Alignment matrix

$$J_{CO}^{(c,d)} = \sum_{w^r \in V_R} \max(0, 1 - f(c^w, d) + f(c^{w^r}, d)) + J_{TEO\text{-}en \to zh} = \|V_{zh} - A_{en \to zh} V_{en}\|^2$$

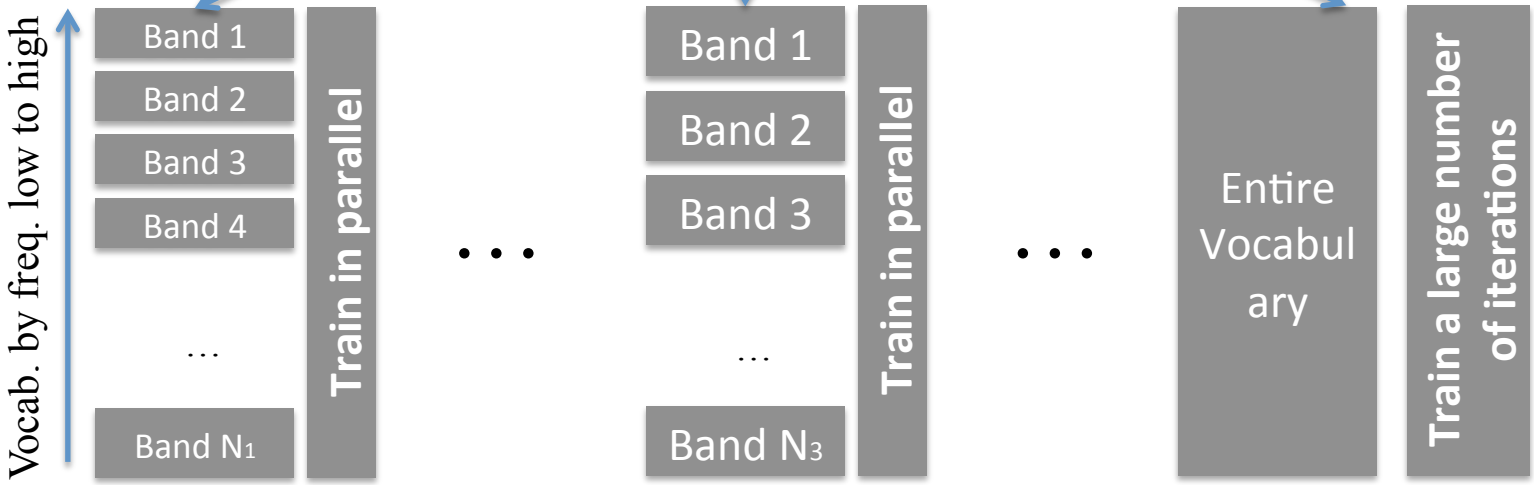
Combining a Neural Language Model with Bilingual Constraints

- Max-margin contrastive objective for learning word embeddings
- Obtain word alignments using the Berkeley Aligner on parallel text
- Combine both objectives to constrain word embeddings for translational equivalence

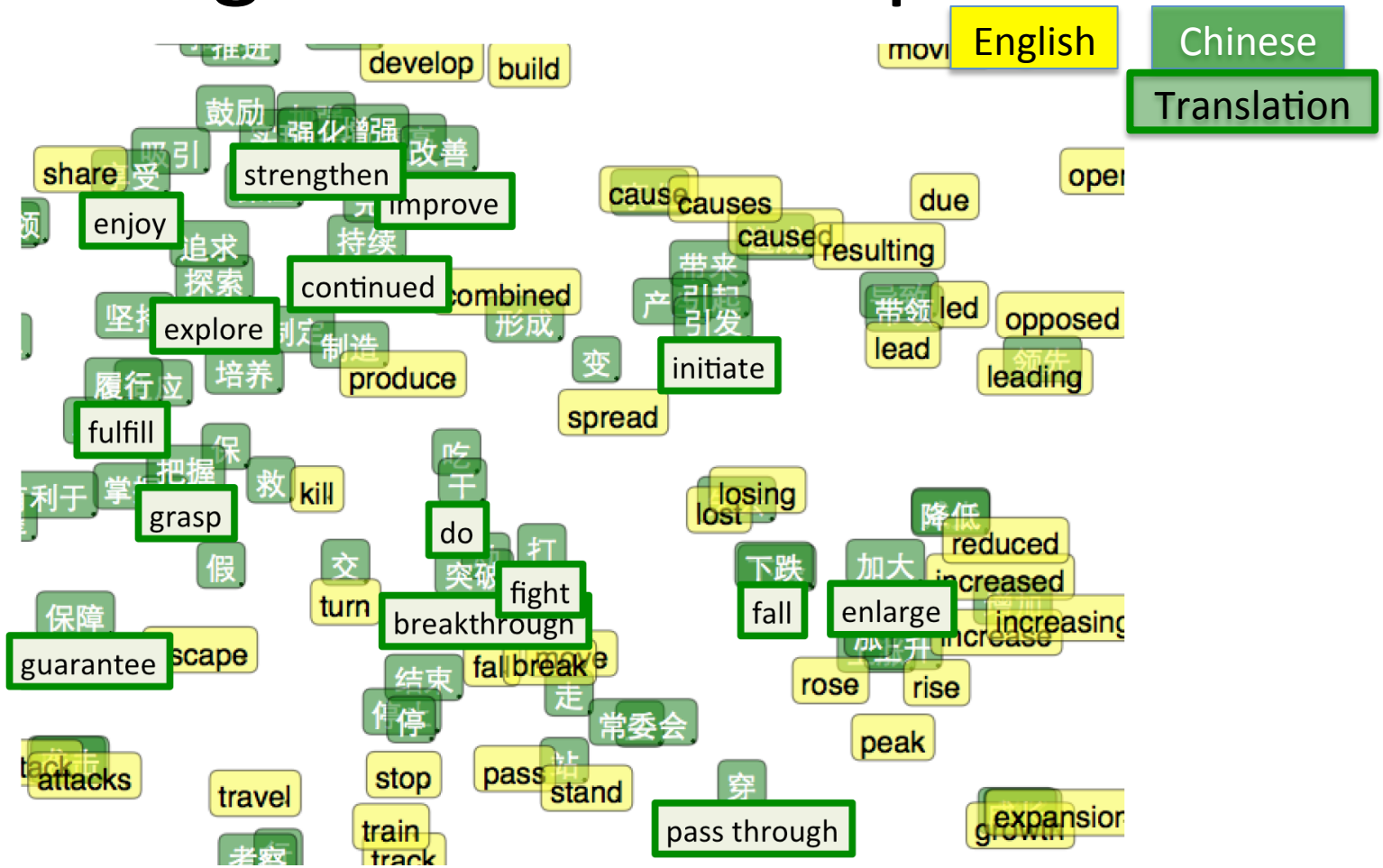
Learning of Embeddings

Band size

Curriculum Training with **minibatch L-BFGS** of varying band sizes:
{5k, 10k, 25k, 50k, 100k}

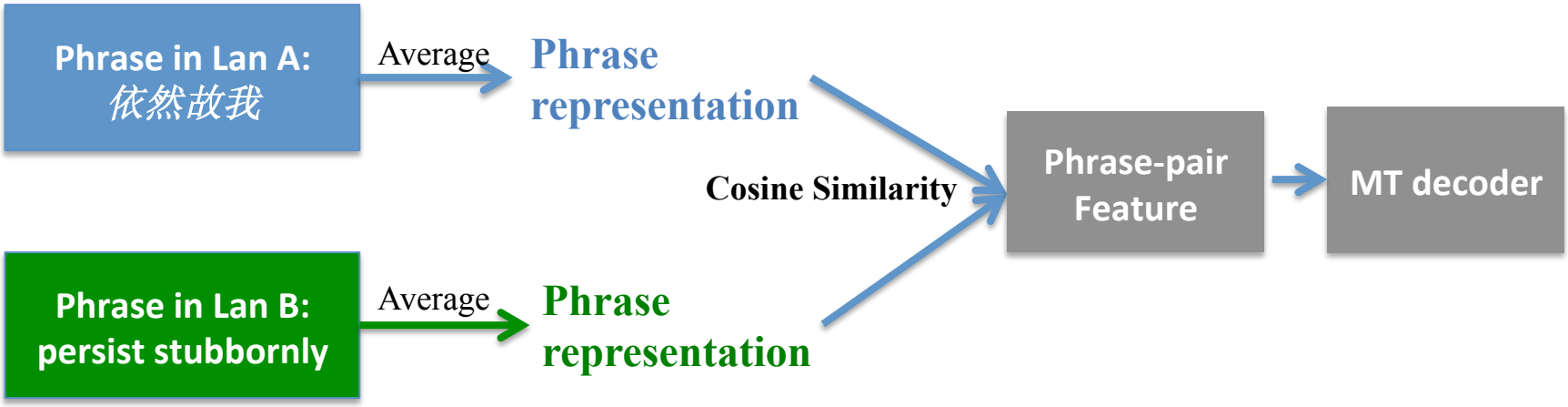


Bilingual Semantic Space



A first set of Mandarin Chinese word embeddings with 100k vocabulary (downloadable from http://ai.stanford.edu/~wzou/mt/)

Application to Stanford Phrasal System



- Phrase-table scoring in an end-to-end MT system
- Competitive BLUE baseline on NIST08 (30.01), with addition data for phrase-table extraction
- Simply average word embeddings to obtain phrase representations
- Cosine similarity is used to form an MT feature
- MERT for decoder optimization

Main Results

Word semantic similarity on SemEval 2012

NIST08 Chinese-English machine translation

Word semantic similarity

Method	*Sp. Corr.* (×100)	*K. Tau* (×100)
Prior work (Jin and Wu, 2012)		5.0
Tf-idf		
Naive tf-idf	41.5	28.7
Pruned tf-idf	46.7	32.3
Word Embeddings		
Align-Init	52.9	37.6
Mono-trained	59.3	42.1
Biling-trained	**60.8**	**43.3**

BLEU score on NIST08 Chinese-English translation task

Method	*BLEU*
Our baseline	30.01
Embeddings	
Random-Init Mono-trained	30.09
Align-Init	30.31
Mono-trained	30.40
Biling-trained	**30.49**