Normalization of Language Embeddings for Cross-Lingual Alignment

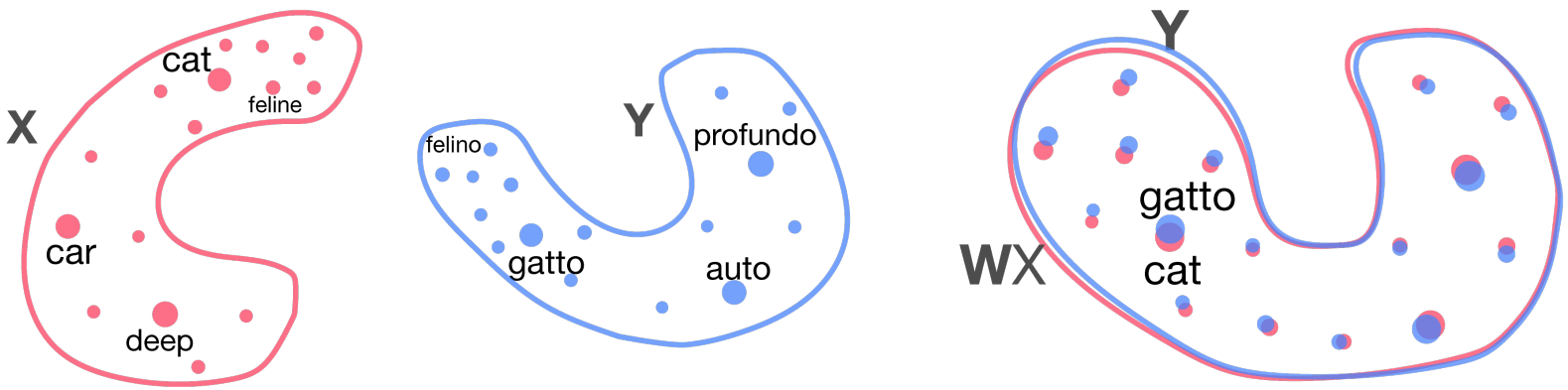
Prince Osei Aboagye[1] Yan Zheng[2] Chin-Chia Michael Yeh[2]
Junpeng Wang[2] Wei Zhang[2] Liang Wang[2] Hao Yang[2]
Jeff M. Phillips[1]

[1]University of Utah

[2]Visa Research

April 15, 2022

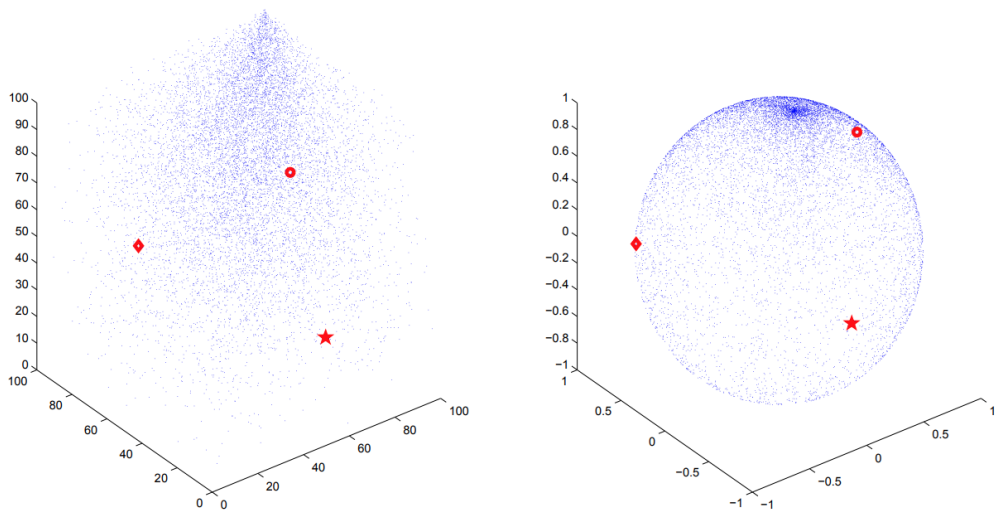
Why Cross-Lingual NLP?



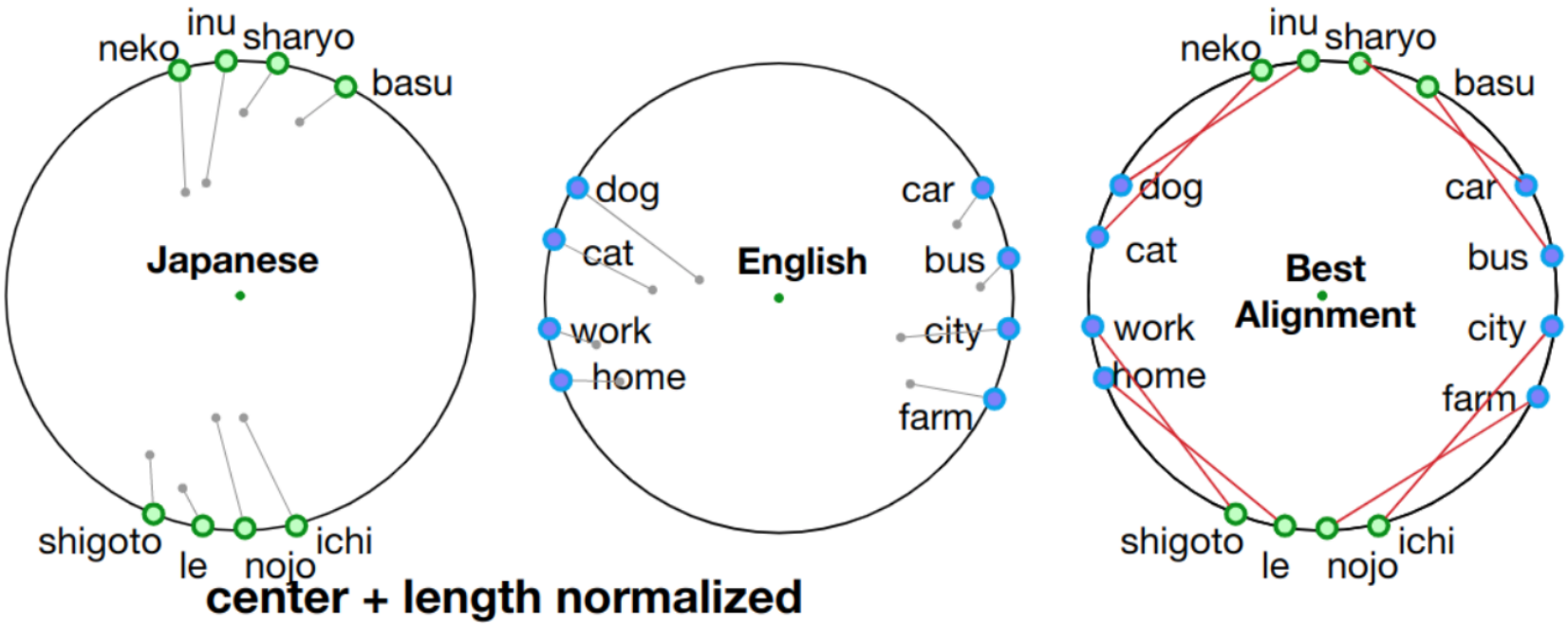
Conneau, et al., 2017

Preprocessing or Normalization Methods

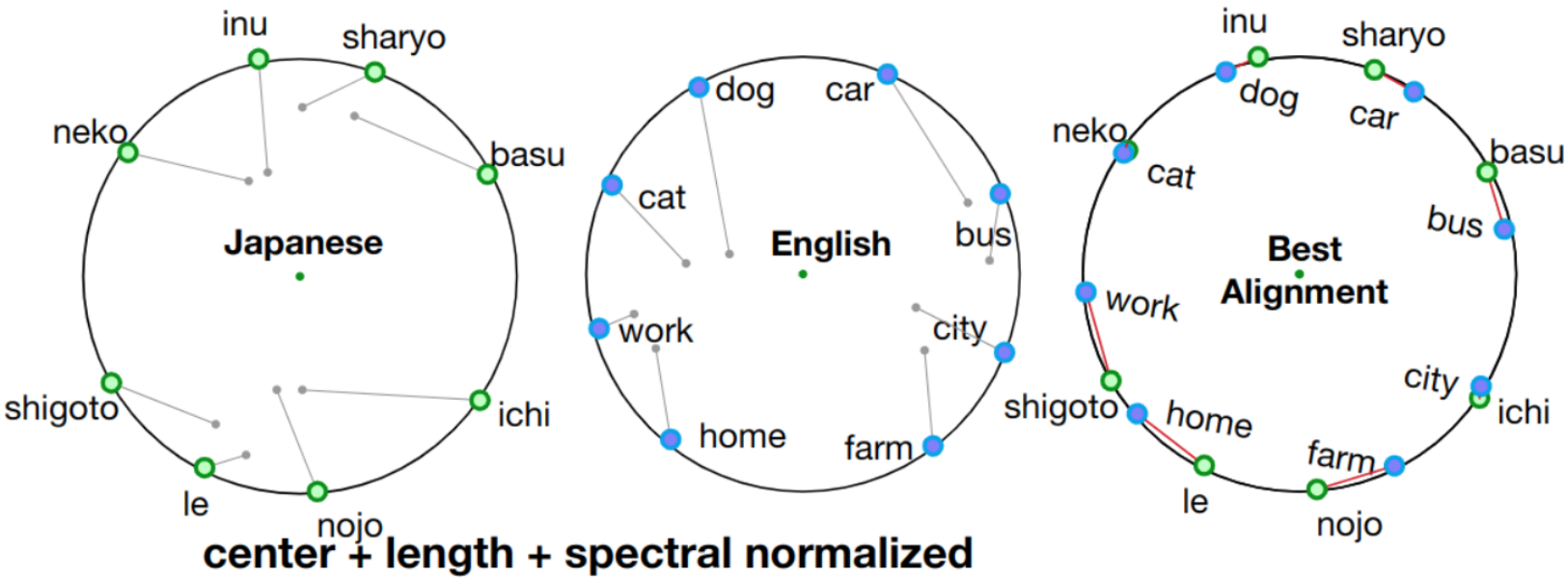
- Mean Centering (C), (*Artetxe et al., 2016*)
- Length Normalization (L), (*Xing et al., 2015*)
- Iterative Normalization (I-C+L), (*Zhang et al., 2019*)



Spectral Normalization (SN)

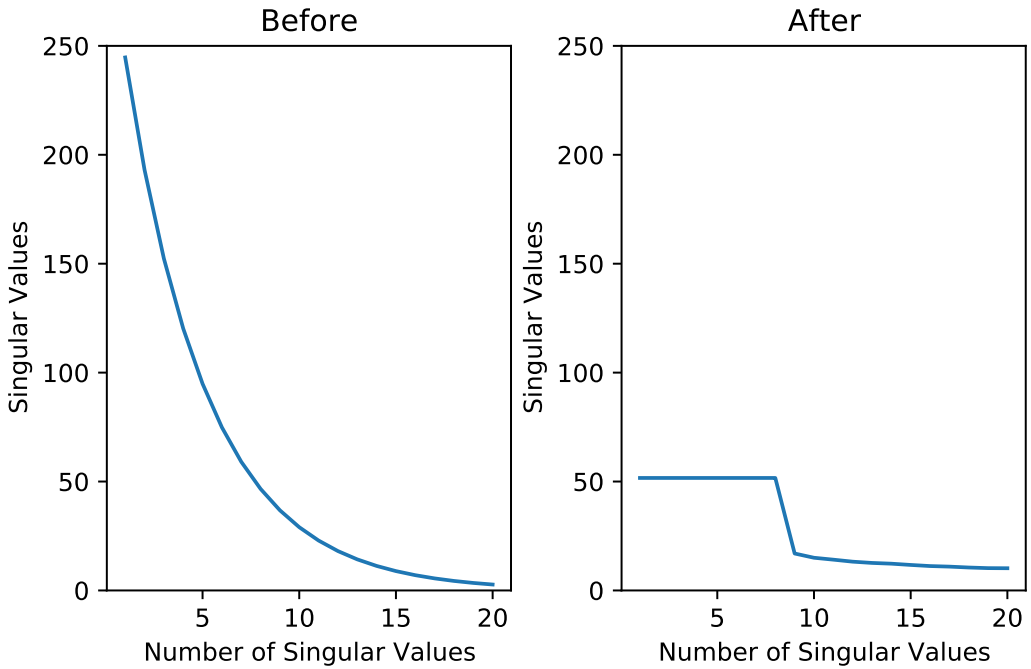


Spectral Normalization (SN)



Spectral Normalization (SN)

- Given a matrix, A of n words in d dimensions.
- Compute the Singular Values.
- Even out the top ones and leave the bottom ones below.



Iterative Spectral Normalization (I-C+SN+L) Algorithm

Algorithm 1 I-C+SN+L

for #Iter steps **do**
    $A \leftarrow$ Center A
    $A \leftarrow$ SpecNorm(A)
    $A \leftarrow$ Unit length normalization of A
return A

Blingual Lexicon Induction (BLI) Task

- Retrieve the nearest neighbors in the Target Language.

Language pairs where I-C+SN+L outperforms I-C+L

Models					
CCA	PROC	PROC-B	DLV	RCSLS	VECMAP
22/28	**24/28**	**25/28**	**23/28**	14/28	**24/28**

Cross-lingual Natural Language Inference (XNLI)

XNLI Performance (Test Set Accuracy)

Model	Dict	EN-DE	EN-FR	EN-TR	EN-RU	Avg
PROC	5k	.607	.534	.568	.585	.574
PROC$^{\text{I-C+L}}$	5k	.589	.608	.536	.581	.579
PROC$^{\text{I-C+SN+L}}$	5k	.611	.638	.542	.596	**.597**
PROC-B	3k	.615	.532	.573	.599	.580
PROC-B$^{\text{I-C+L}}$	5k	.602	.636	.537	.595	.593
PROC-B$^{\text{I-C+SN+L}}$	3k	.624	.638	.548	.601	**.603**
RCSLS	5k	.390	.363	.387	.399	.385
RCSLS$^{\text{I-C+L}}$	5k	.514	.490	.490	.526	.505
RCSLS$^{\text{I-C+SN+L}}$	5k	.499	.482	.504	.556	**.510**

- We introduce a new way to normalize embeddings based on **Spectral Normalization**.
- Our approach:
  - ⋆ Generalizes previous approaches, and
  - ⋆ Effectively removes much of the clustering of words.
  - ⋆ Allows any alignment procedure to find better alignments resulting in improved performance on the BLI and CLDTs.