

# Trained Mappingを用いた 多言語Cross-lingual Word Embedding

九州大学大学院

システム情報科学府 自然言語処理研究室

ハンボムソク

# 概要

## ■ 3つの言語Word Embedding(分散表現)を一つのものに組み合わせる

- 相性の良い言語ペアの既学習Cross-lingual Word Embedding(CWE)を活用し、
- 他の言語へのCWEの性能を改善

例

英語 - スペイン語

日本語

### Cross-lingual Word Embedding(CWE)

#### 既存のCWEモデル

#### Trained Mappingを用いた多言語CWE

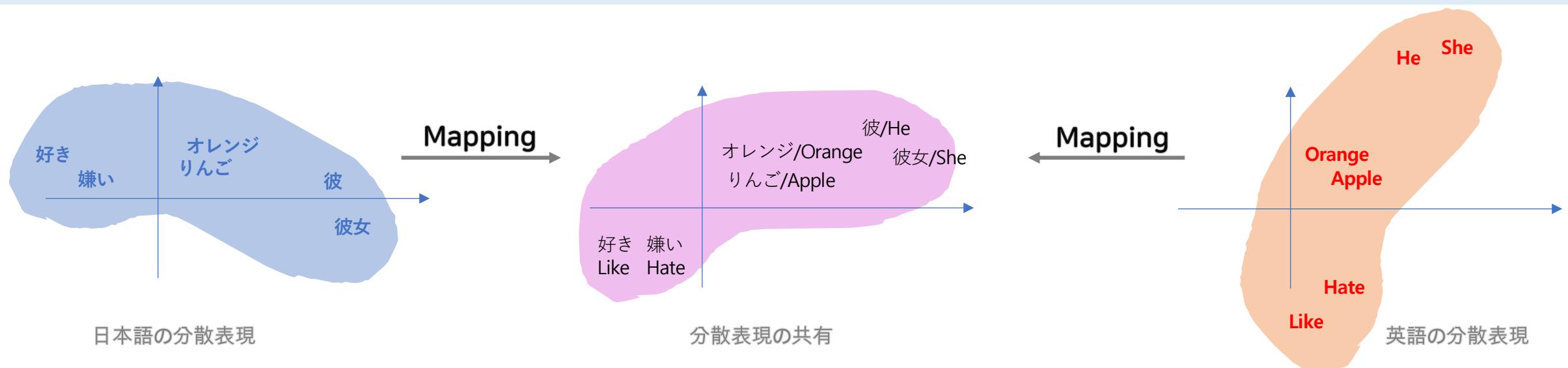
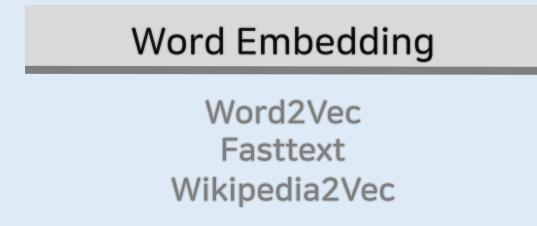
#### 今後の課題

#### 参考資料

# Cross-lingual Word Embedding(CWE)

## ■ Cross-lingual Word Embeddingとは

- 異なる言語分散表現をMappingし、共有させること。
  - 異なる言語の分散表現は違う方式で数値化されている。
  - しかし、言語の中での分布は似ている。

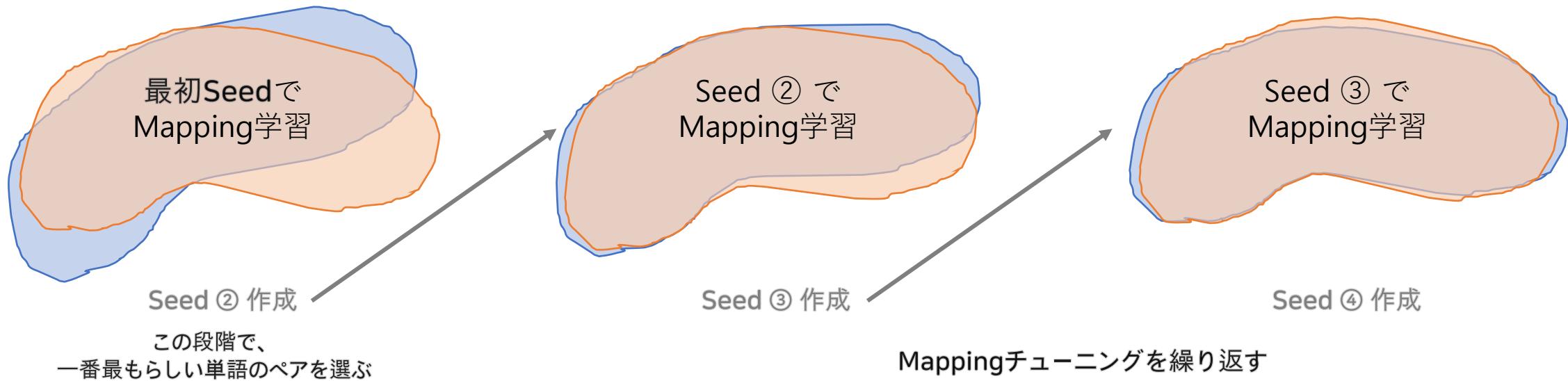


重み(Mapping)を学習させ、分散表現が合わせる。

# Cross-lingual Word Embedding(CWE)

## ■ Cross-lingual Word Embeddingのやり方

最初Seed作成 → Mapping学習 → 新たなSeedの作成 → Mappingのチューニング



## 最初Seedの作成方法により

### Supervised CWE

Dictionary : いくつかの辞書情報を使う

Identical word : 記号、英語単語などの情報を使う

### Unsupervised CWE

敵対的生成ネットワーク(GAN)を用い  
分散表現の分布を合わせる

# Cross-lingual Word Embedding(CWE)

## ■ Cross-lingual Word Embeddingの活用

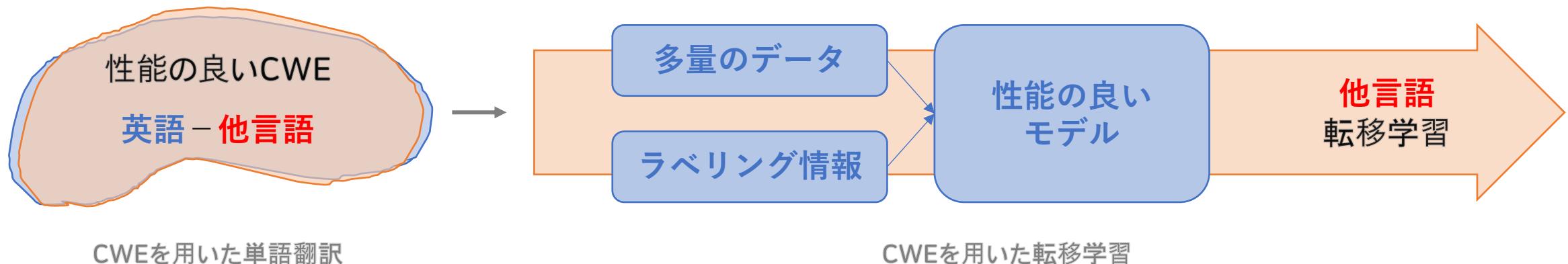
- 単語単位の翻訳

Mapping後の分散表現で最も類似度の高い単語を選ぶことで、単語の翻訳意味を予測する。

- 転移学習

ある学習済みモデルを、新たな領域のモデルとして使用する。

英語 : データ量・ラベリング情報など、学習へのメリットがある言語



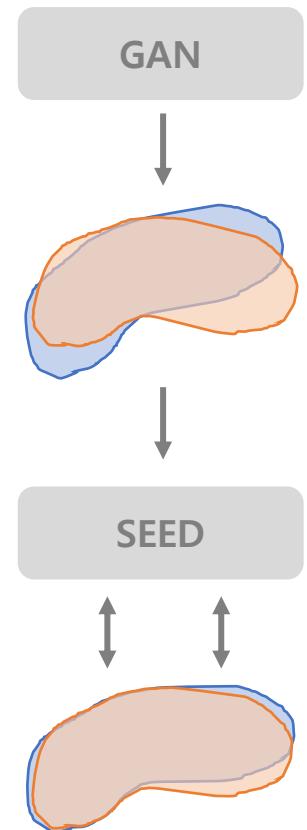
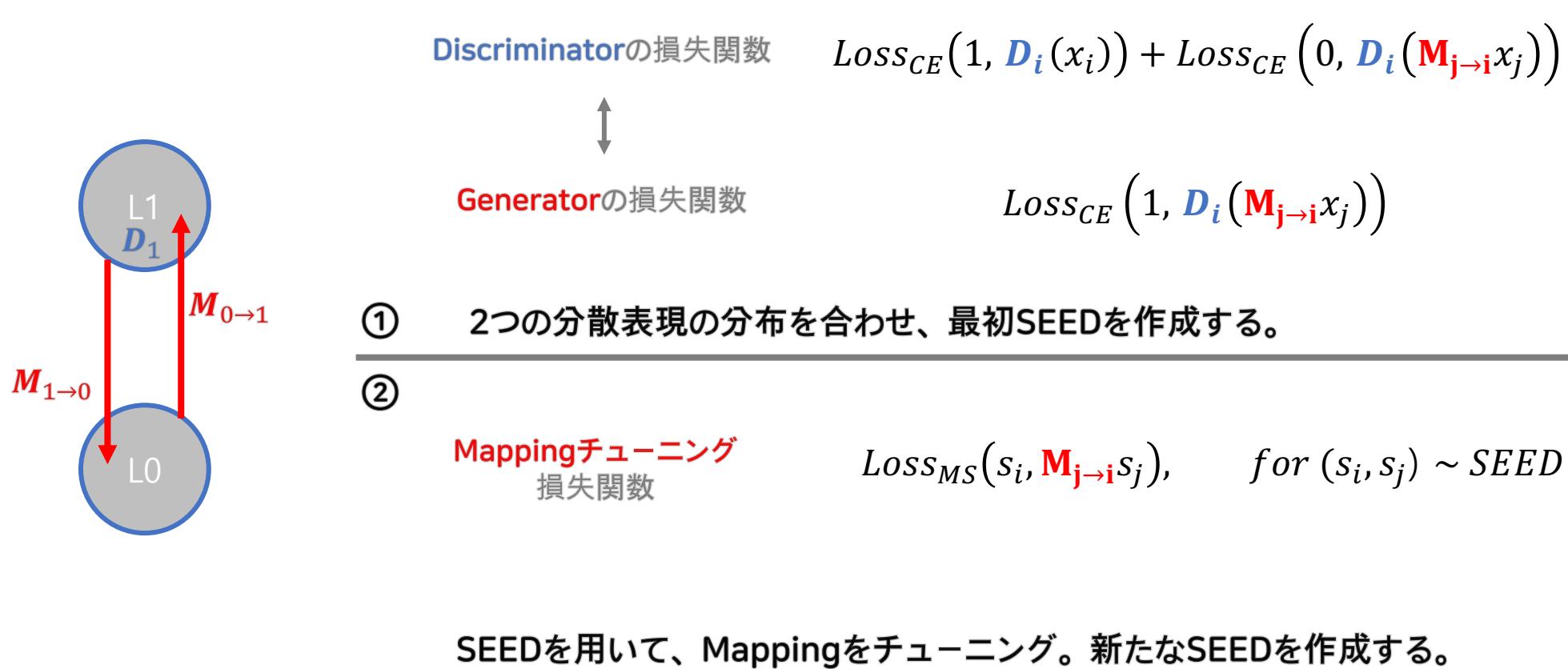
情報の多いResource Rich言語のメリットを共有

# 既存のCWEモデル

## ■ 教師なしCWE(2つの言語)

- ① GANを用いたMappingの学習・最初SEEDの作成
- ② Mappingチューニング

$Loss_{CE}(t, p)$	: Cross-Entropy Loss Function
	: $-[t \log(p) + (1-t) \log(1-p)]$
$Loss_{MS}(t, p)$	: Mean Square Loss Function
	: $\frac{1}{n} \sum (t_i - p_i)^2$
$M_i$	: Mapping from $L_i$ to $L_p$
$D_i$	: Discriminator for $L_i$
$X_i$	: Embedding space of $L_i$
$x_i$	: Embedding vector of $X_i$

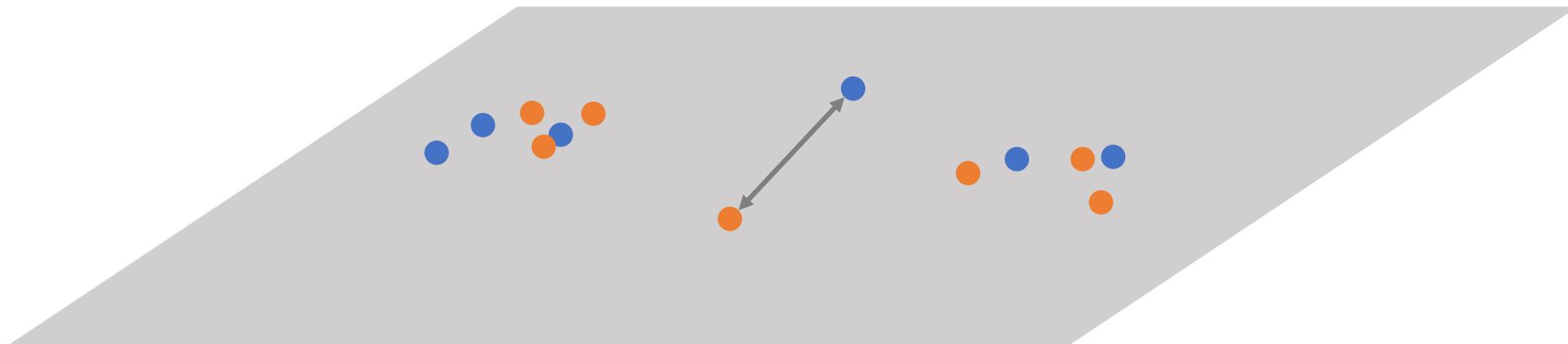


## \* SEEDの作成

### ■ CSLS(Cross-Lingual Similarity Scaling)のMutual Nearest Neighbors関係

- CSLS: 最近傍法、高次元スペースでのhubness問題を解除
- CSLS最小値のペアが、「 $x_i \rightarrow x_j$ 」 and 「 $x_j \rightarrow x_i$ 」になった時、 $(x_i, x_j)$ はSEEDになる。

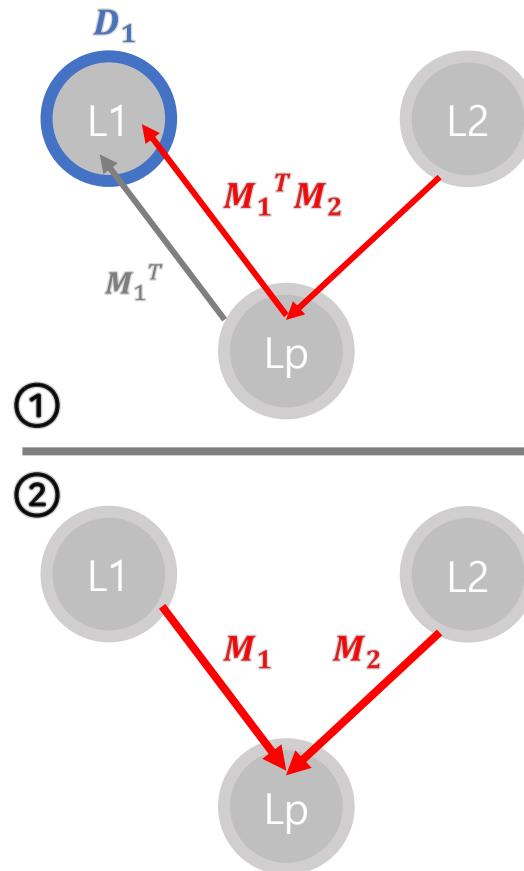
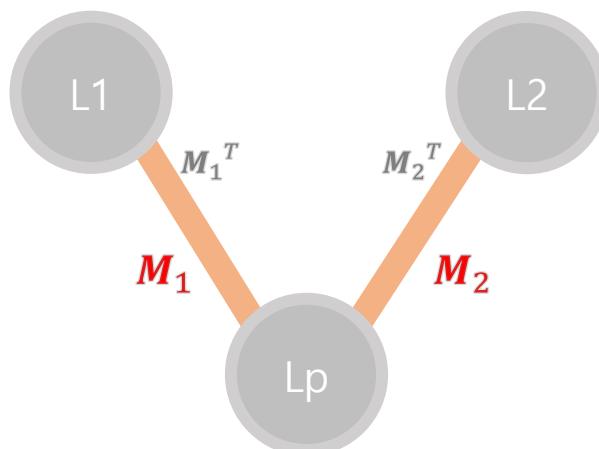
$$CSLS(x, y) = 2 \cos(x, y) - \frac{1}{n} \sum_{y' \in N_Y(x)} \cos(x, y') - \frac{1}{n} \sum_{x' \in N_X(y)} \cos(x', y)$$



# 既存のCWEモデル

## 教師なし多言語CWE

- Pivot言語( $L_p$ )に複数の言語をMapping
- ランダムに2つの言語を選択し、Mappingを学習



$Loss_{CE}(t, p)$ : Cross-Entropy Loss Function : $[-t \log(p) + (1-t) \log(1-p)]$
$Loss_{MS}(t, p)$ : Mean Square Loss Function : $\frac{1}{n} \sum (t_i - p_i)^2$
$M_i$ : Mapping from $L_i$ to $L_p$
$D_i$ : Discriminator for $L_i$
$X_i$ : Embedding space of $L_i$
$x_i$ : Embedding vector of $X_i$

$i, j \in \{pivot, 1, 2\}$

$$Loss_{CE}(1, D_i(x_i)) + Loss_{CE}(0, D_i(\mathbf{M}_i \mathbf{M}_j^T x_j))$$

GAN

$$Loss_{CE}(1, D_i(\mathbf{M}_i \mathbf{M}_j^T x_j))$$

2言語の場合より、Mappingのヴァリエーションが増える。

---


$$Loss_{MS}(s_p, M_i s_i), \quad \text{for } (s_p, s_i) \sim SEED$$

*from  $(X_p, M_i X_i)$*

$$Loss_{MS}(\mathbf{M}_i s_i, \mathbf{M}_j s_j), \quad \text{for } (s_i, s_j) \sim SEED$$

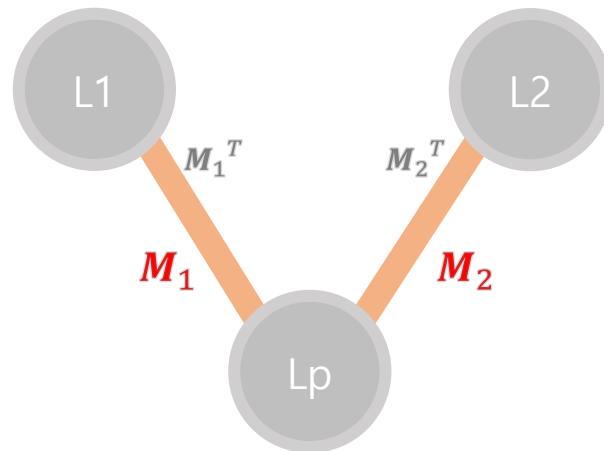
*from  $(M_i X_i, M_j X_j)$*

$L_i, L_j$ 間のSEEDを用いMappingチューニングができる。

# 既存のCWEモデル

## ■ 教師なし多言語CWEの弱点

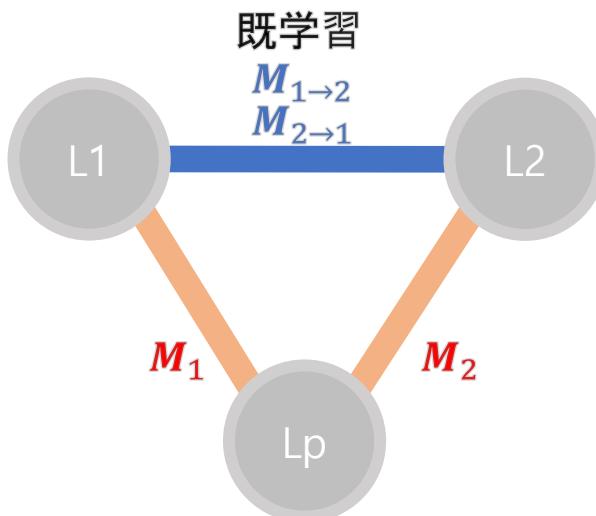
- Pivot言語 ( $L_p$ ) と相性の良い(分散表現の類似度が高い)言語( $L_1, L_2, \dots$ )が望ましい。
- $L_i, L_j$ 間のSEEDを作成する時、GANで学習されたMapping性能に頼る。



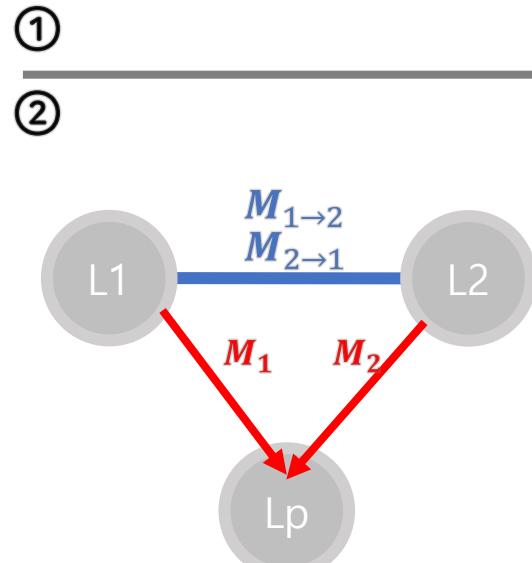
Pivot言語と相性の悪い言語ペアの学習が不安定である。

# Trained Mappingを用いた多言語CWE

- 性能の良いTrained Mappingを使用することで、良質なSEEDの作成が可能



教師なし多言語CWE  
と  
同様に行う。



$$Loss_{CE}(1, D_i(x_i)) + Loss_{CE}(0, D_i(M_i M_j^T x_j)) \quad i, j \in \{pivot, 1, 2\}$$

GAN

$$Loss_{CE}(1, D_i(M_i M_j^T x_j))$$

---

$$Loss_{MS}(s_p, \mathbf{M}_1 s_1), \quad \text{for } (s_p, s_1) \sim SEED$$

from  $(X_p, \mathbf{M}_1 X_1)$

$$from (\mathbf{X}_p, \mathbf{M}_2 \mathbf{M}_{1 \rightarrow 2} X_1)$$
  
  
$$Loss_{MS}(\mathbf{M}_1 s_1, \mathbf{M}_2 s_2), \quad \text{for } (s_1, s_2) \sim SEED$$

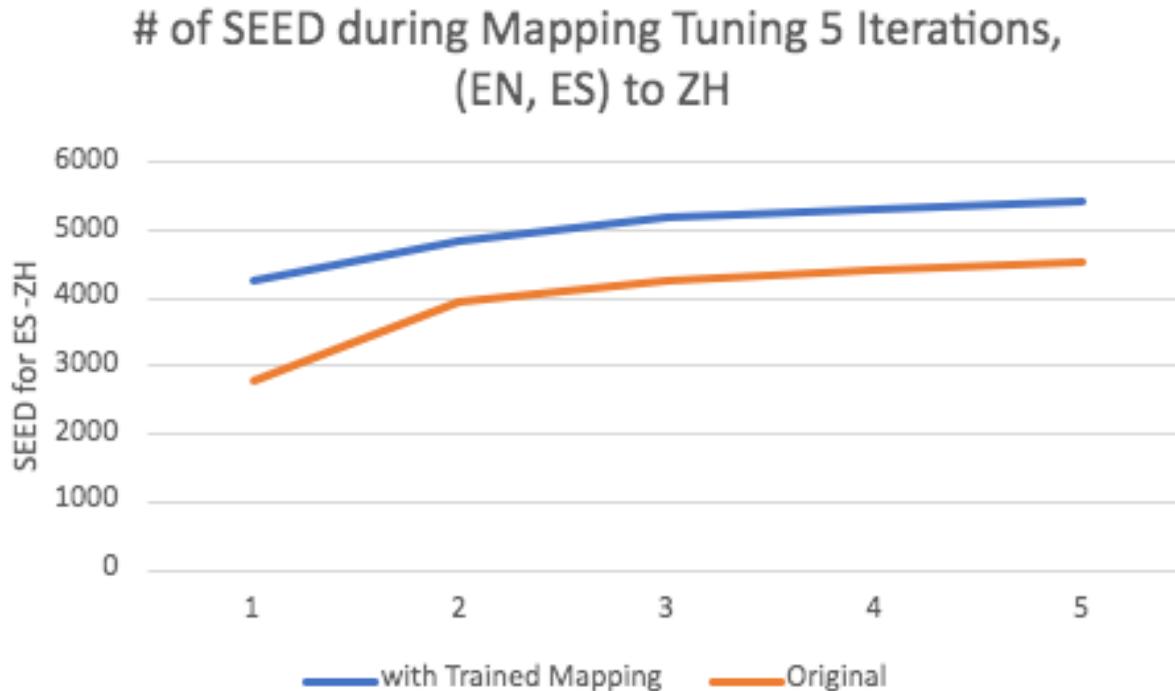
from  $(\mathbf{M}_1 X_1, \mathbf{M}_1 X_1)$

from  $(X_1, \mathbf{M}_{2 \rightarrow 1} X_2)$

from  $(X_2, \mathbf{M}_{1 \rightarrow 2} X_1)$

# Trained Mappingを用いた多言語CWE

## 実験結果①：MappingチューニングでのSEEDが増える。



Average # of SEED (上:Original, 下:Trained Map)

		(EN, ES)		
		ZH	PL	JA
3882	↓	3789	↓	4223
4854		5561		5365

### 実験条件

Word Embedding : Wikipedia2Vec

Size of Embeddings : 100,000

EN-ES Mapping性能 : 83%

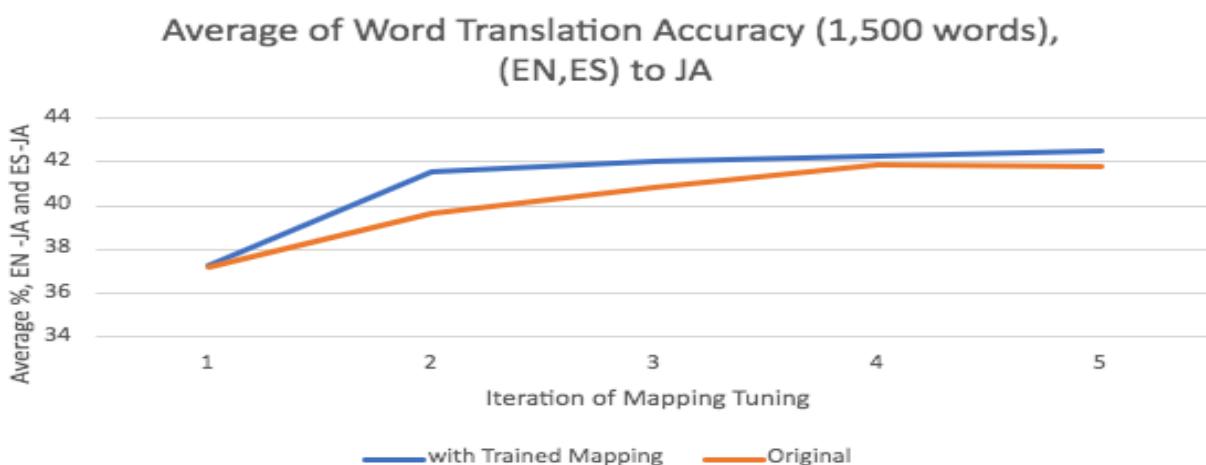
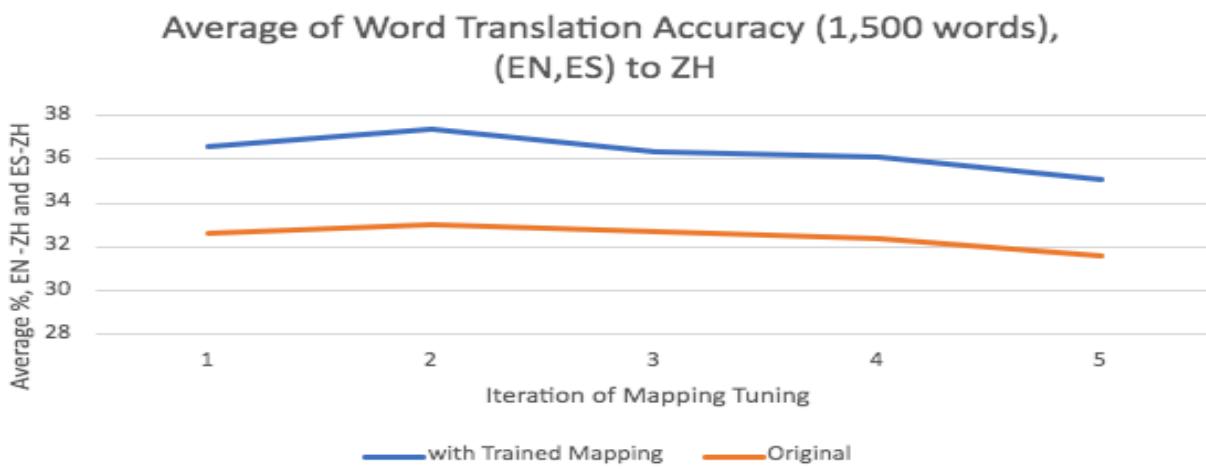
ES-EN Mapping性能 : 81%

pre-trained(GAN) mappingは同じ

# Trained Mappingを用いた多言語CWE

## 実験結果②：Mappingチューニングの性能が上がる。

- 特に分散表現分布の類似度が低い言語に関してのCWE性能



Max Performance of Word Translation  
(上:Original, 下:Trained Map)

EN (%)			ES (%)		
ZH	PL	JA	ZH	PL	JA
38.3	58.2	48.0	28.1	42.7	36.2
42.0	59.9	49.4	33.0	42.3	35.7

### 実験条件

Word Embedding : Wikipedia2Vec  
Size of Embeddings : 100,000  
EN-ES Mapping性能 : 83%  
ES-EN Mapping性能 : 81%

MUSE Dataset(Facebook)  
1500個の辞書情報を用いた評価

# Trained Mappingを用いた多言語CWE

## ■ 実験結果③：誤ったGAN Trained Mappingを改善する。

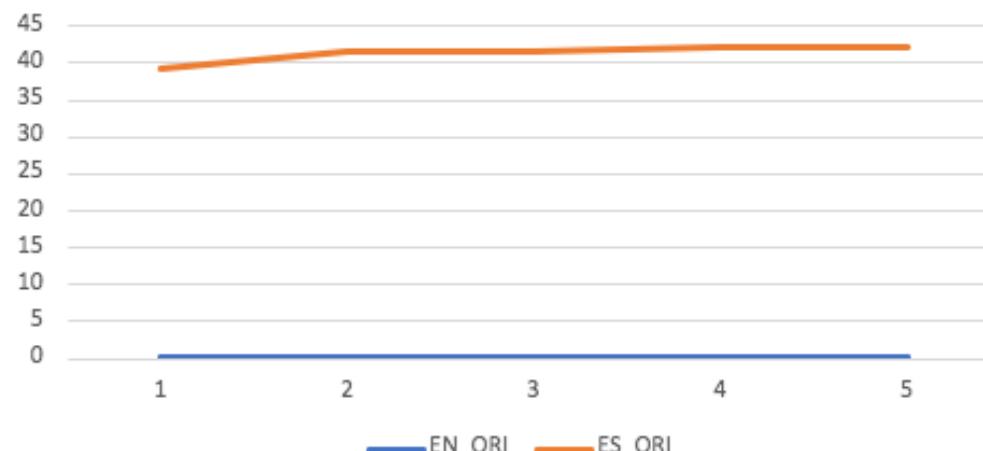
- 場合により、GANでの学習が上手く進まないことがある。

### 誤ったGANの Pre Trained Mapping (EN, ES) - PL

EN - PL での性能が、ほぼ0%の状態でチューニングに進む。

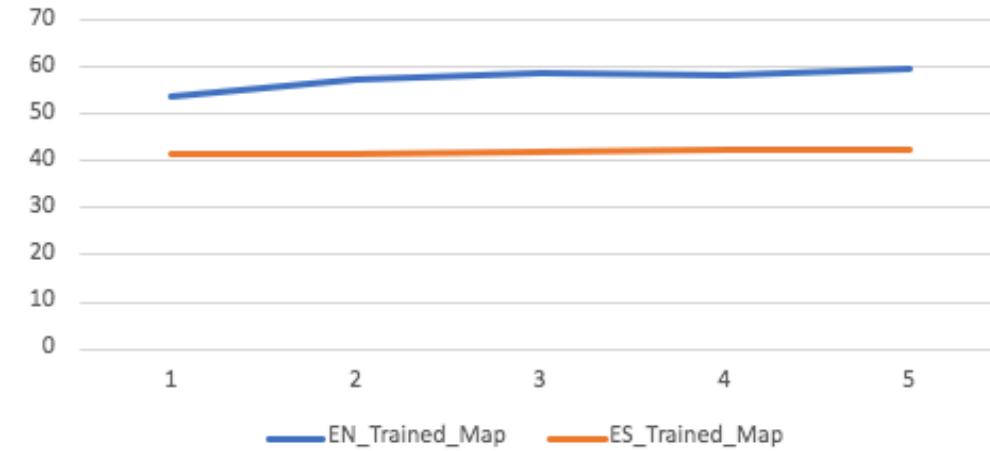
#### Original Mapping Tuning

Word Translation Accuracy (Original Model),  
(EN, ES) to PL



#### Mapping Tuning using Trained Mapping

Word Translation Accuracy (Trained Mapping),  
(EN, ES) to PL



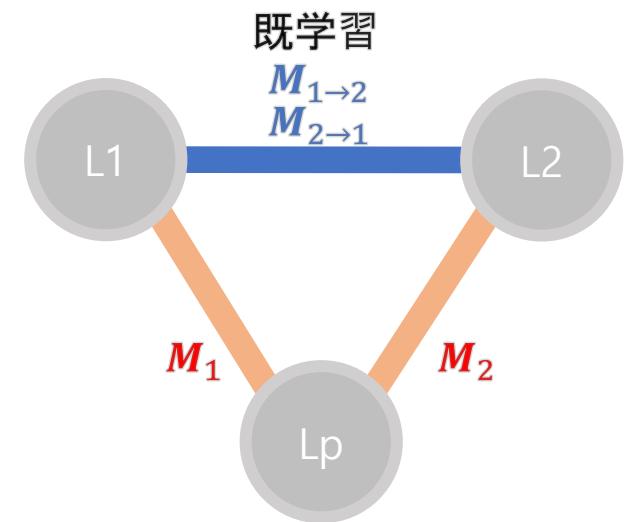
# Trained Mappingを用いた多言語CWE

## ■ まとめ

- 既存モデルの弱点 (UMWE)
  - Pivot言語に頼ったSEEDの作成
  - 相性の悪い言語ペア (分散表現分布の類似度: 低) の学習性能
- 提案: Pivot言語とは関係のない「GOOD Trained Mapping」を用い、SEED作成・Pivot言語へのMappingを改善
- 結果: 良質なSEEDが得られ、  
Pivot言語との相性が悪い時、Mappingの性能が同上する。  
Pre-trained Mappingに誤りがある時、Mappingチューニング学習を直す。

**Create SEED for  $L_i, L_p$**   
*from  $(X_p, M_1 X_1)$*   
*from  $(X_p, M_2 M_{1 \rightarrow 2} X_1)$*

**Create SEED for  $L_i, L_j$**   
*from  $(M_1 X_1, M_1 X_1)$*   
*from  $(X_1, M_{2 \rightarrow 1} X_2)$*   
*from  $(X_2, M_{1 \rightarrow 2} X_1)$*



# 今後の課題

## ■ 実験方法の多様化

- 入力言語の多様化

- Low-Resource言語に対する実験を行う。

- 言語の相性による改善モデルの効果を具体化する。

- 評価方法の多様化

- 大規模データセットにおいてのWord Translationテストを行う。

- CWEを用いた転移学習（翻訳・感情分析・文章分類など）での性能評価を行う。

Low-Resource言語に対する、転移学習の性能改善を最後の目標にする。

# 参考文献

- Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Herv Jgou. 2018. Word translation without parallel data. In International Conference on Learning Representations.
- Xilun Chen, Claire Cardie. 2018. Unsupervised Multilingual Word Embeddings, Proceedings of the 2018 EMNLP.
- Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, Yuji Matsumoto. 2020. Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia. Proceedings of the 2020 EMNLP.
- Mozhi Zhang, Keyulu Xu, Ken-ichi Kawarabayashi, Stefanie Jegelka, Jordan Boyd-Graber. 2019. Are Girls Neko or Shōjo? Cross-Lingual Alignment of Non-Isomorphic Embeddings with Iterative Normalization. Proceedings of the 57th Annual Meeting of the ACL.

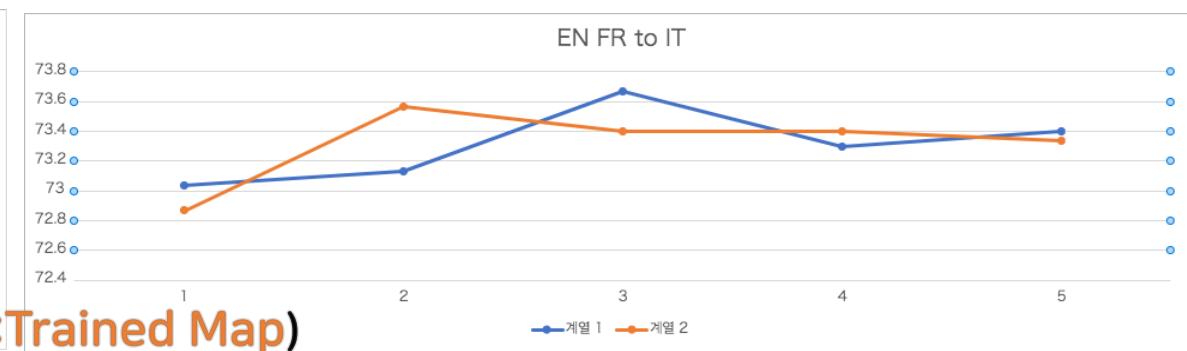
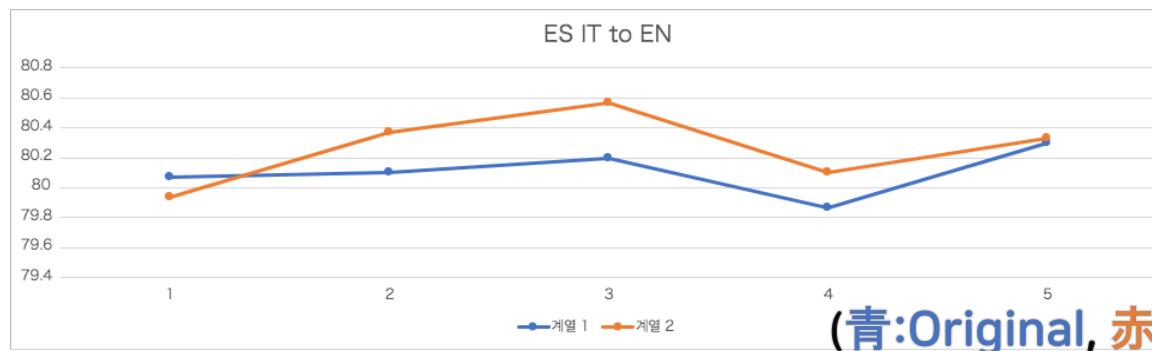
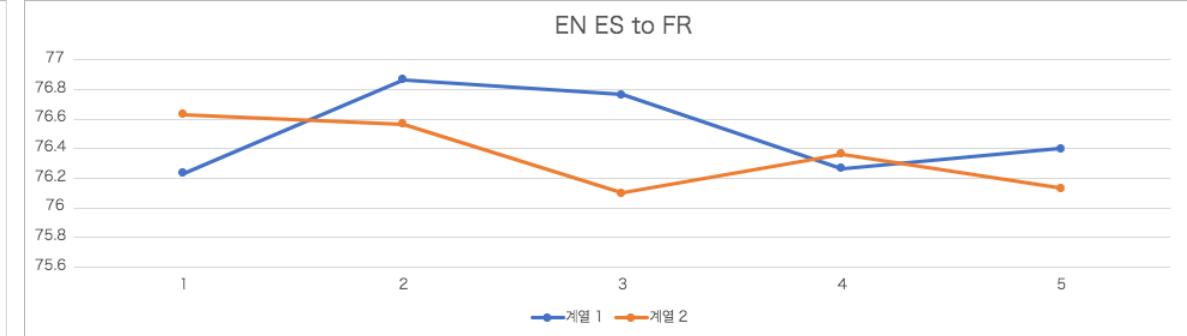
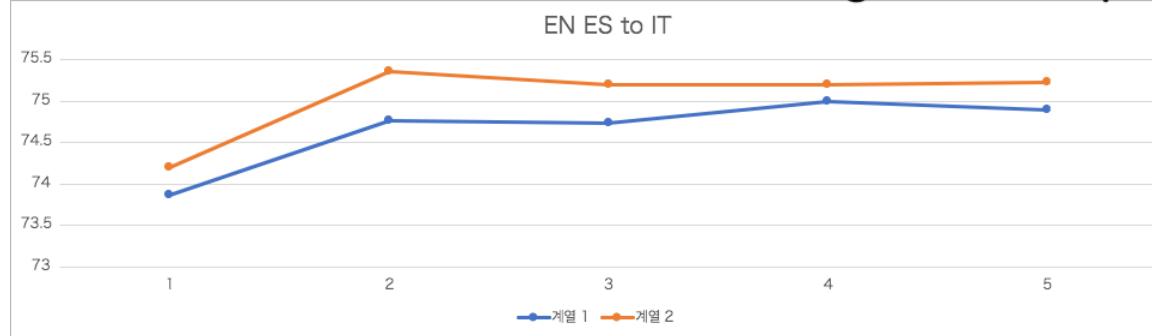
# 参考資料 (相性の良い言語を使用した場合)

## ■ Unsupervised Multilingual Word Embedding(UMWE)モデルとの比較

- Word Embedding(Fasttext), Embedding Size(200,000), Languages(EN, ES, FR, IT, DE)

	es-en	it-en	en-it	fr-it	de-it	en-es	it-es	es-it	en-fr	es-fr	en-de	it-de
paper	83.7	77.7	78.8	83.5	72.0	82.5	88.2	84.5	82.4	86.9	74.8	69.5
UMWE	<b>83.2</b>	77.4	78.4	68.7	69.3	81.9	75.7	71.6	82.4	<b>71.3</b>	74.1	68.4
UMWE_H	<b>83.1</b>	<b>77.7</b>	<b>78.5</b>	<b>68.8</b>	<b>70.5</b>	<b>81.9</b>	<b>76.0</b>	<b>71.9</b>	<b>82.5</b>	<b>70.7</b>	<b>74.5</b>	<b>68.8</b>

Average Accuracy of Word Translation



(青:Original, 赤:Trained Map)

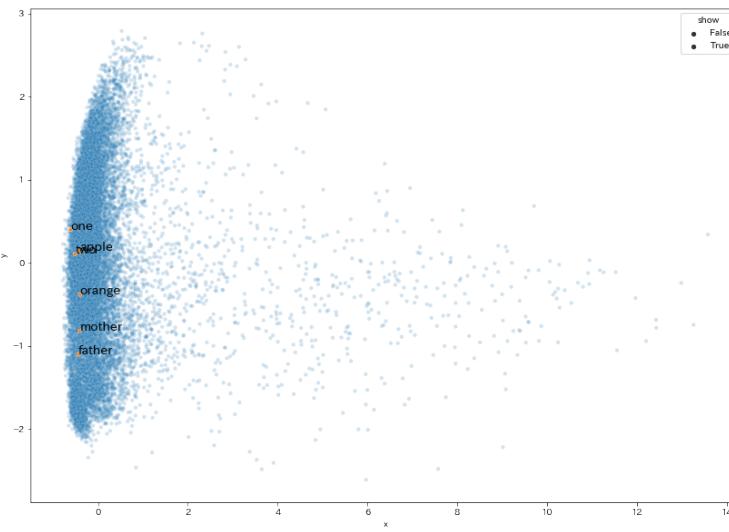
# 参考資料 (Wikipedia2Vecを使った理由①)

## ■ 日本語の分散表現

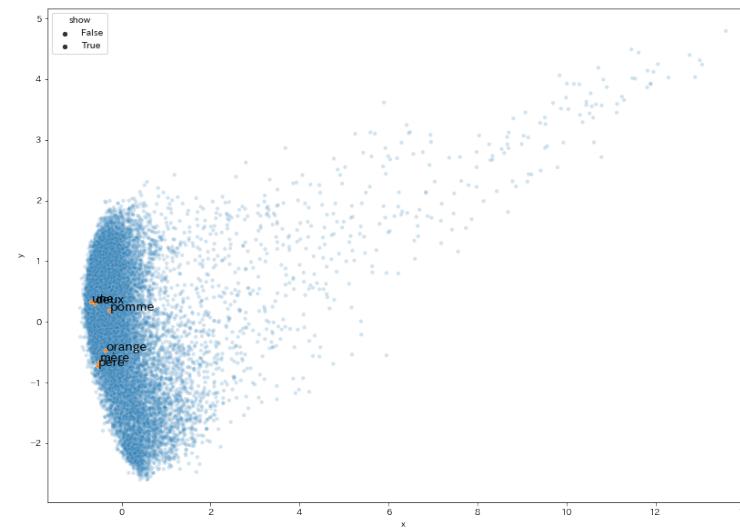
- fasttext

facebookから提供され、157言語の既学習Word Embeddingが得られる

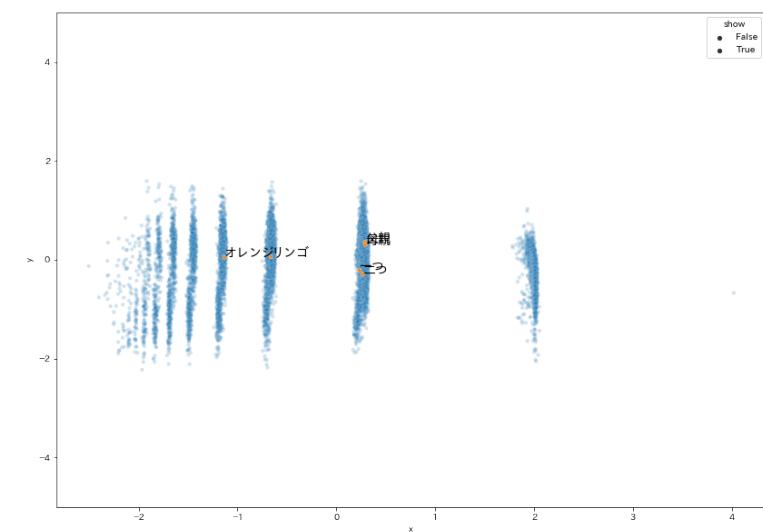
ほとんどのCWE論文で使用され、評価の基準として扱われている



英語の分散表現



フランス語の分散表現

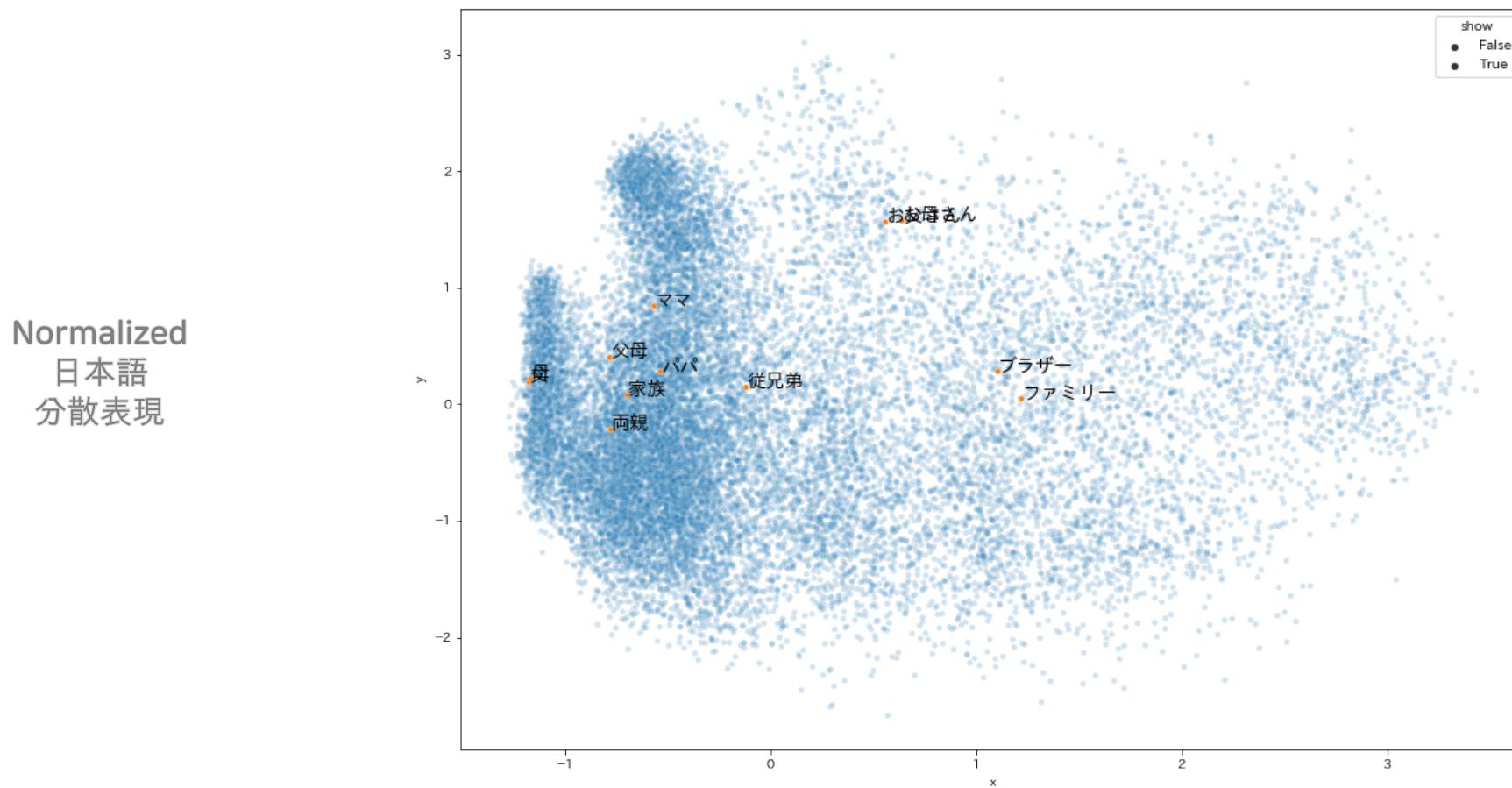


日本語の分散表現

# 参考資料 (Wikipedia2Vecを使った理由②)

## ■ 日本語の分散表現

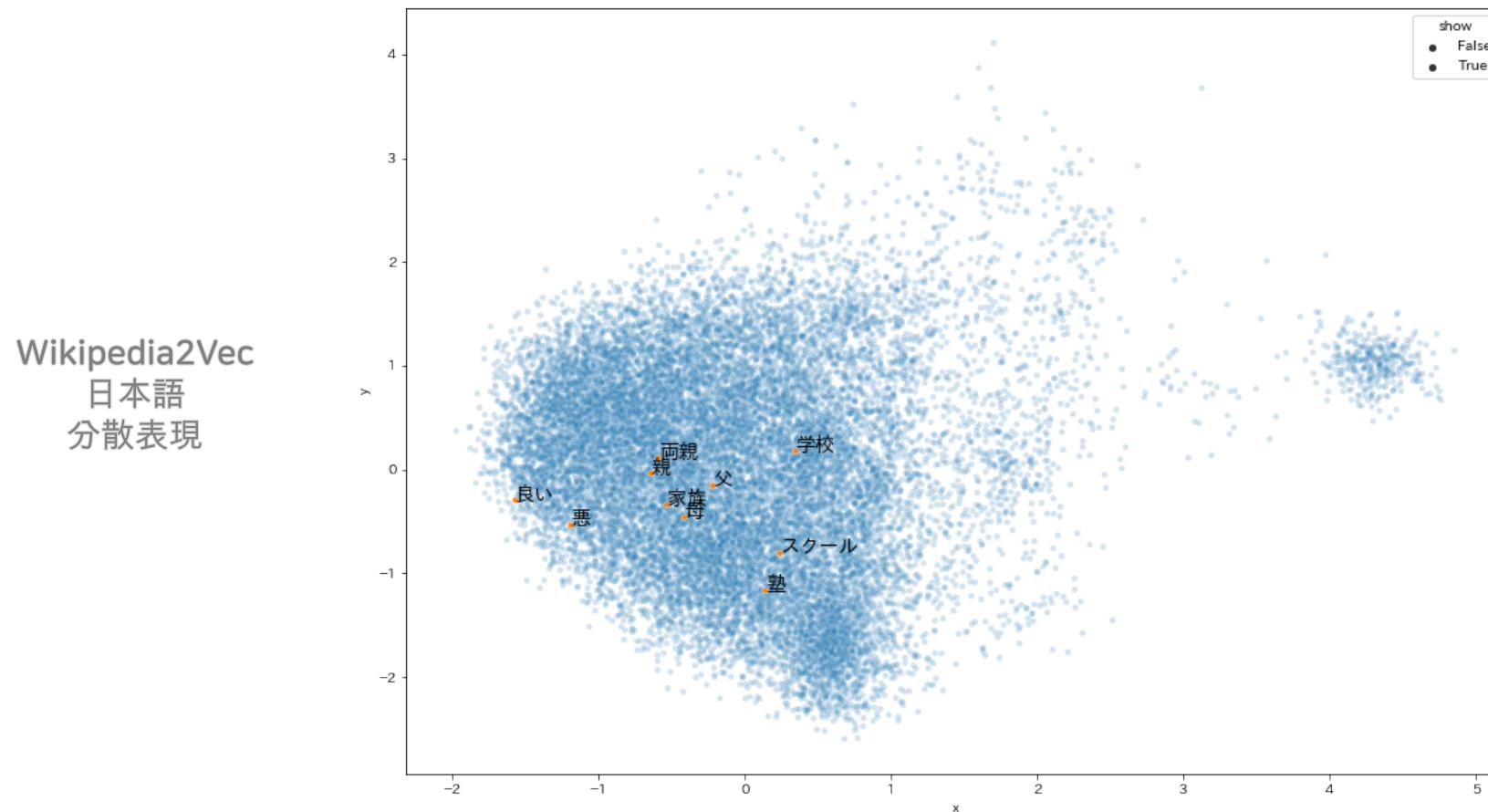
- Normalizing (分布の平均を0、各ベクトルの長さを1にする)



# 参考資料 (Wikipedia2Vecを使った理由③)

## ■ 日本語の分散表現

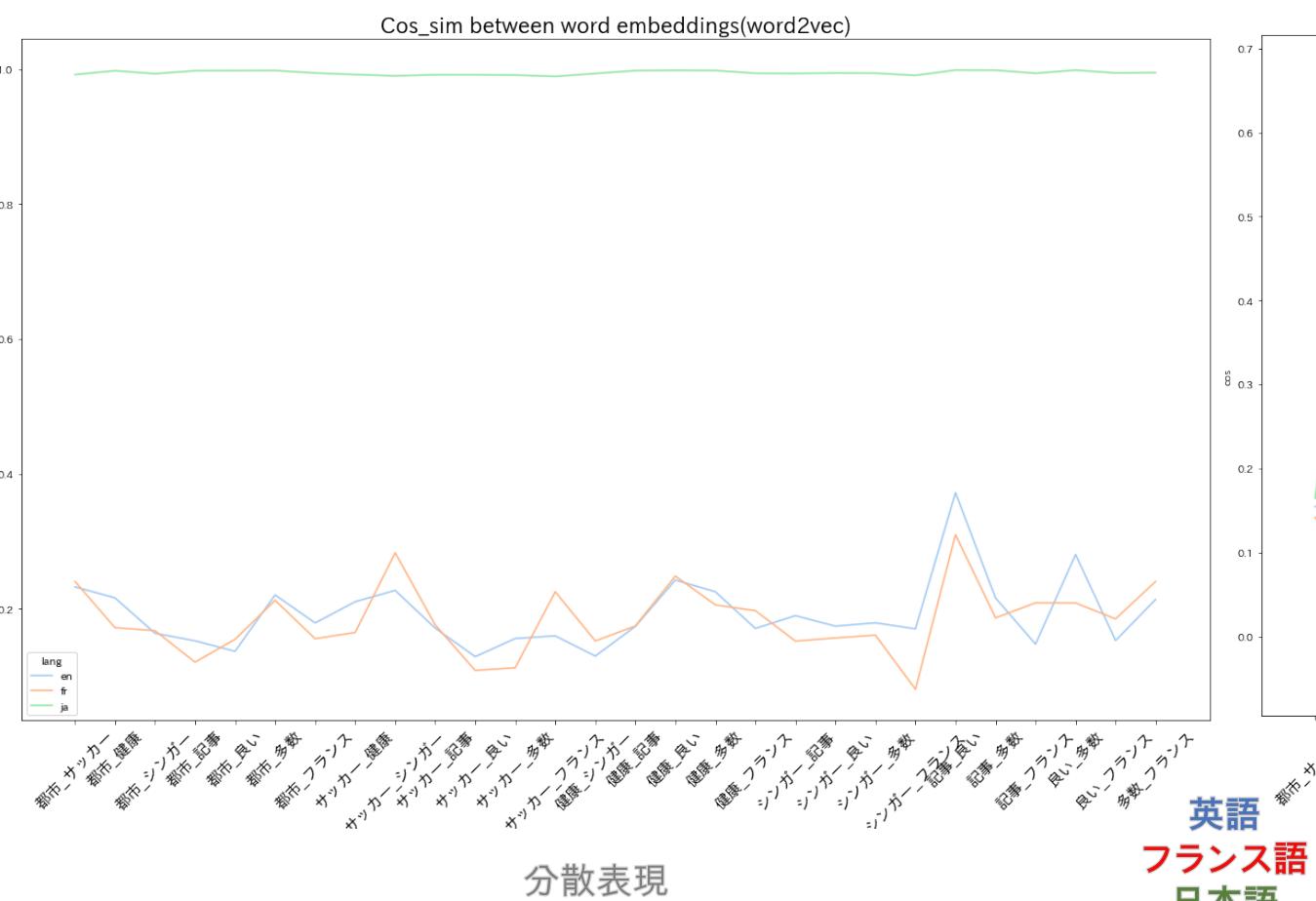
- Wikipedia2Vec (Wikipediaのリンク情報をWord Embeddingの学習に使う)



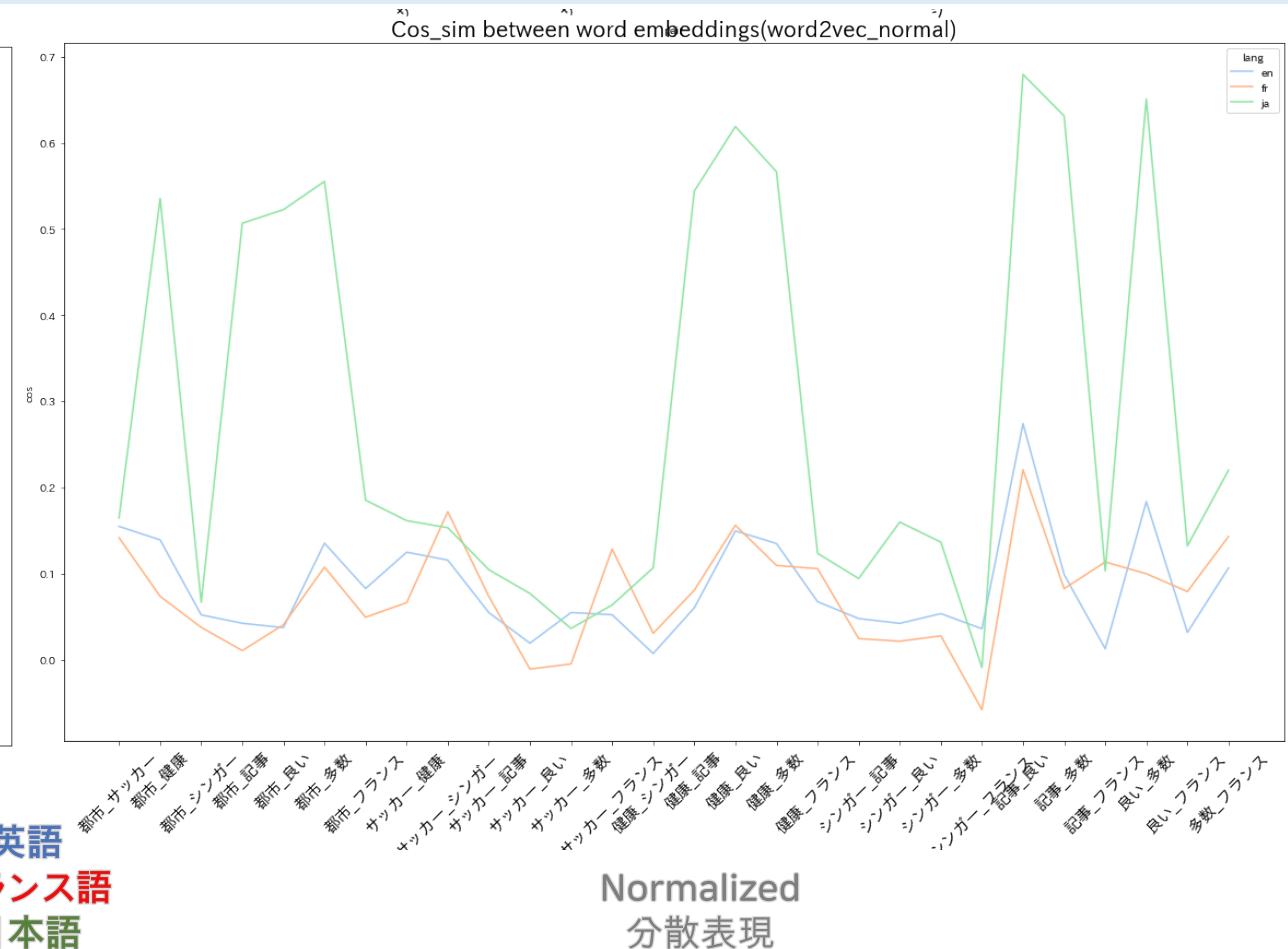
# 参考資料 (Wikipedia2Vecを使った理由④)

## ■ 日本語の分散表現

- ある単語ペアのCosine類似度を、英語・フランス語・日本語の分散表現から計算したグラフ



英語  
フランス語  
日本語



Normalized  
分散表現

# 参考資料 (Wikipedia2Vecを使った理由⑤)

## ■ 日本語の分散表現

- Wikipedia2Vec (Wikipediaのリンク情報をWord Embeddingの学習に使う)

