A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings

Mikel Artetxe, Gorka Labaka, Eneko Agirre

IXA NLP group – University of the Basque Country (UPV/EHU)

Introduction

Introduction

Monolingual corpus

Introduction

Basque

Monolingual corpus

Introduction

Basque

Monolingual corpus

Monolingual corpus

Introduction

Basque — Monolingual corpus

English — Monolingual corpus

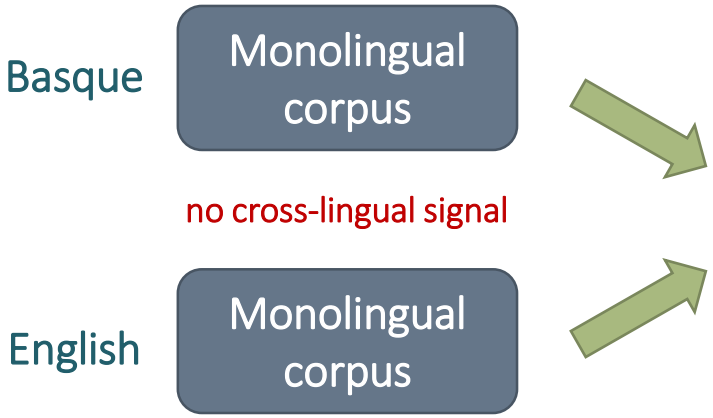
Introduction

Basque — Monolingual corpus

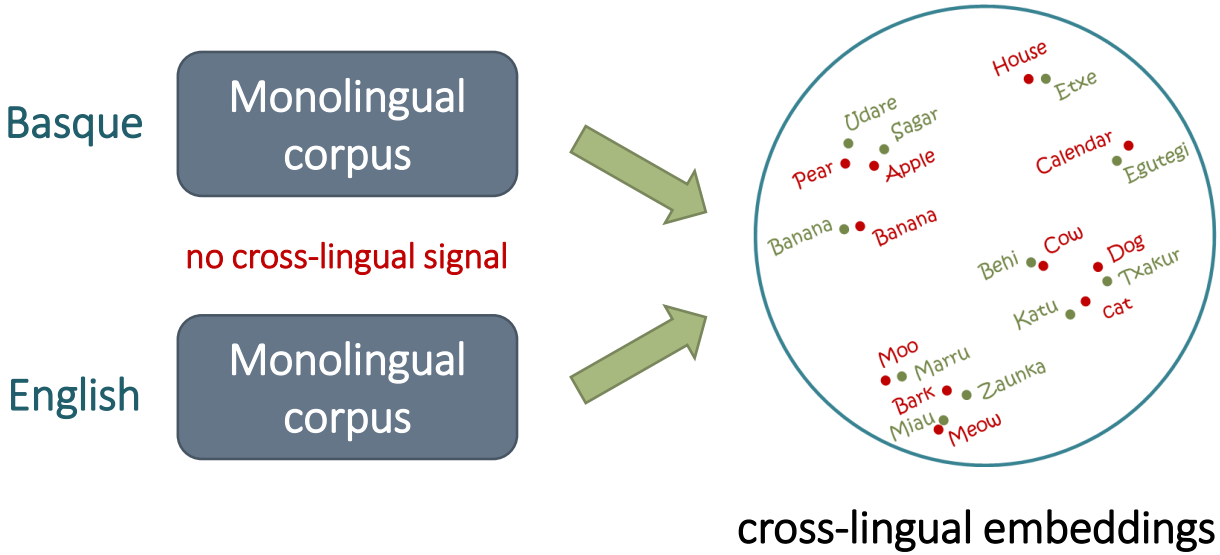
no cross-lingual signal

English — Monolingual corpus

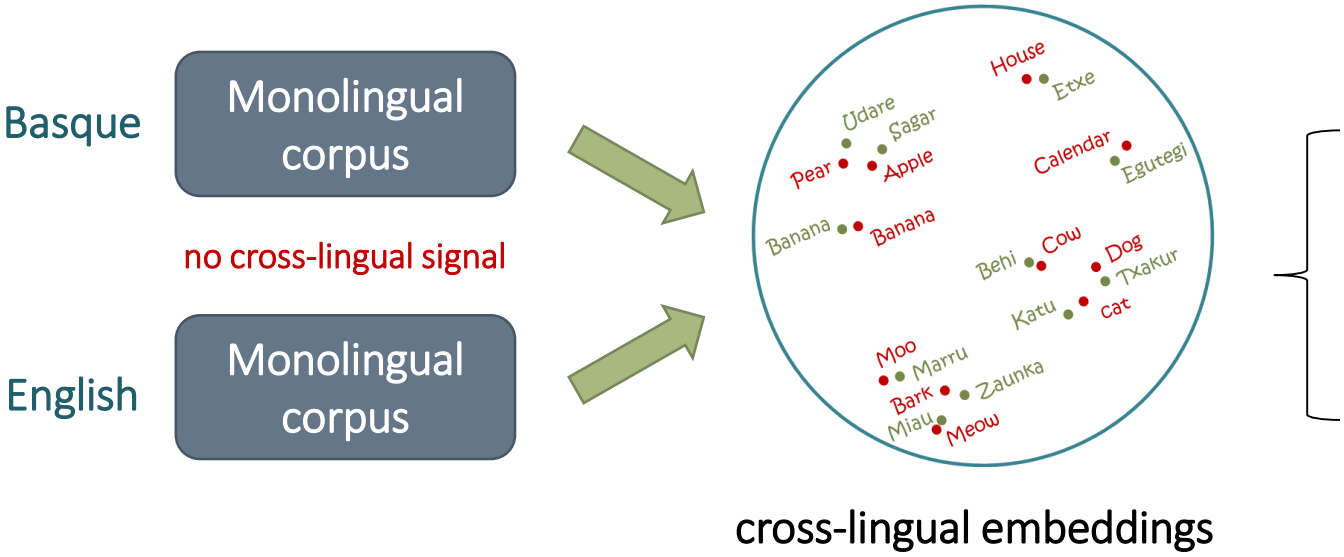
Introduction



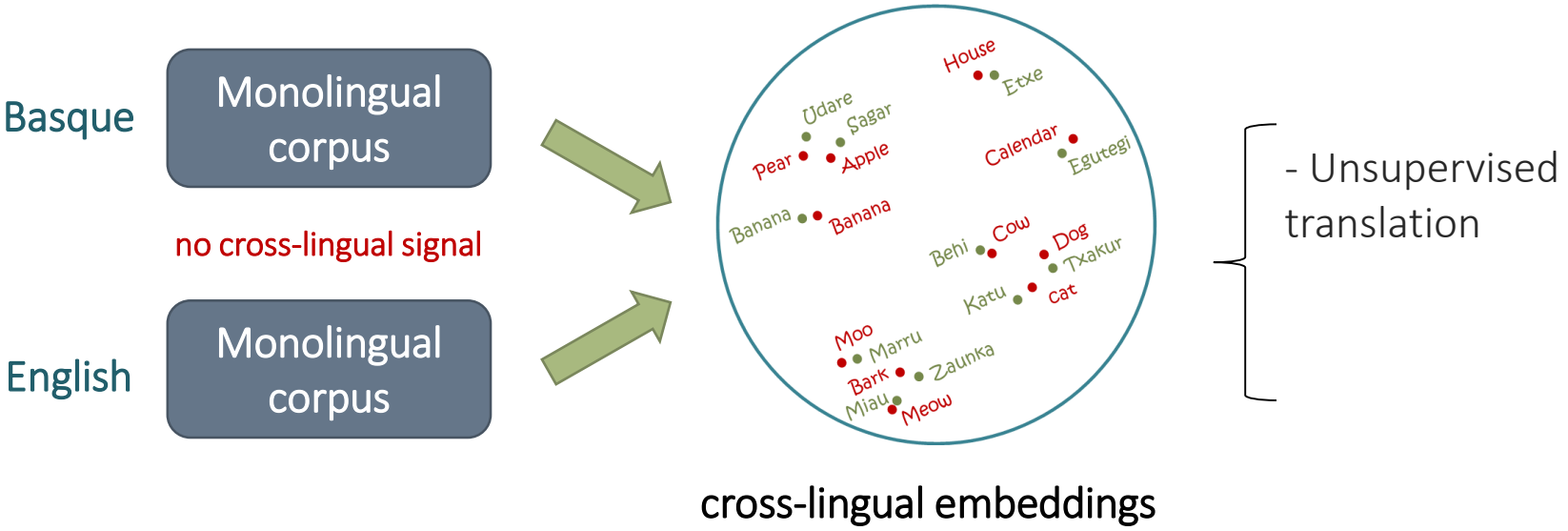
Introduction



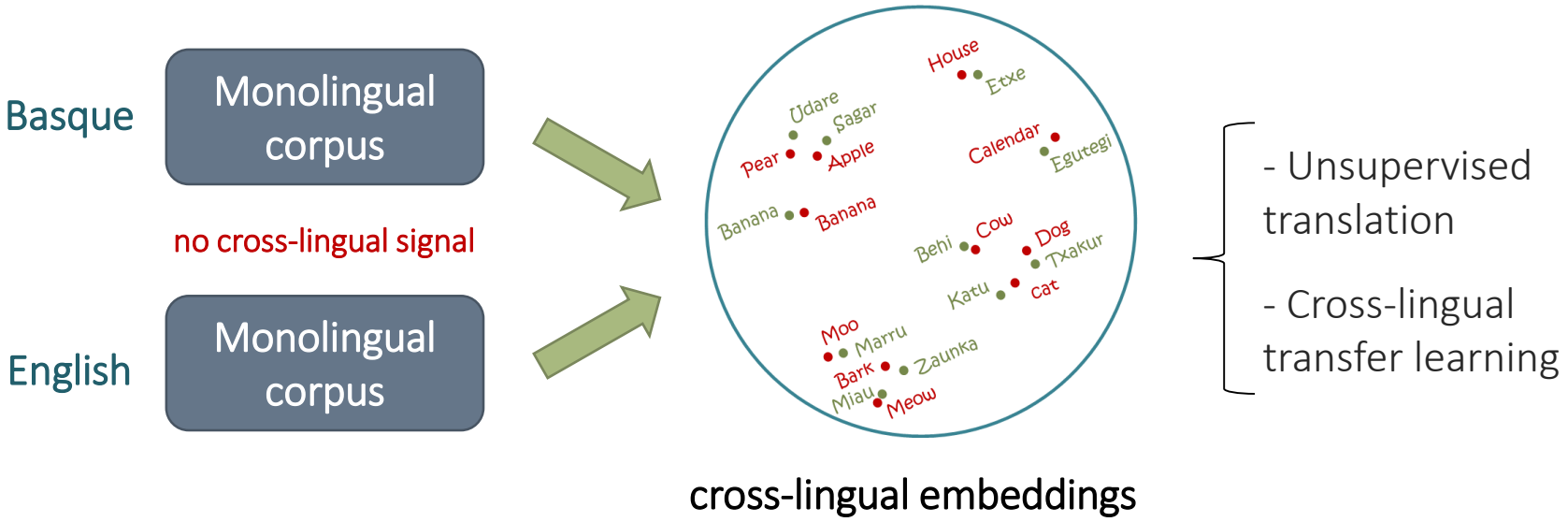
Introduction



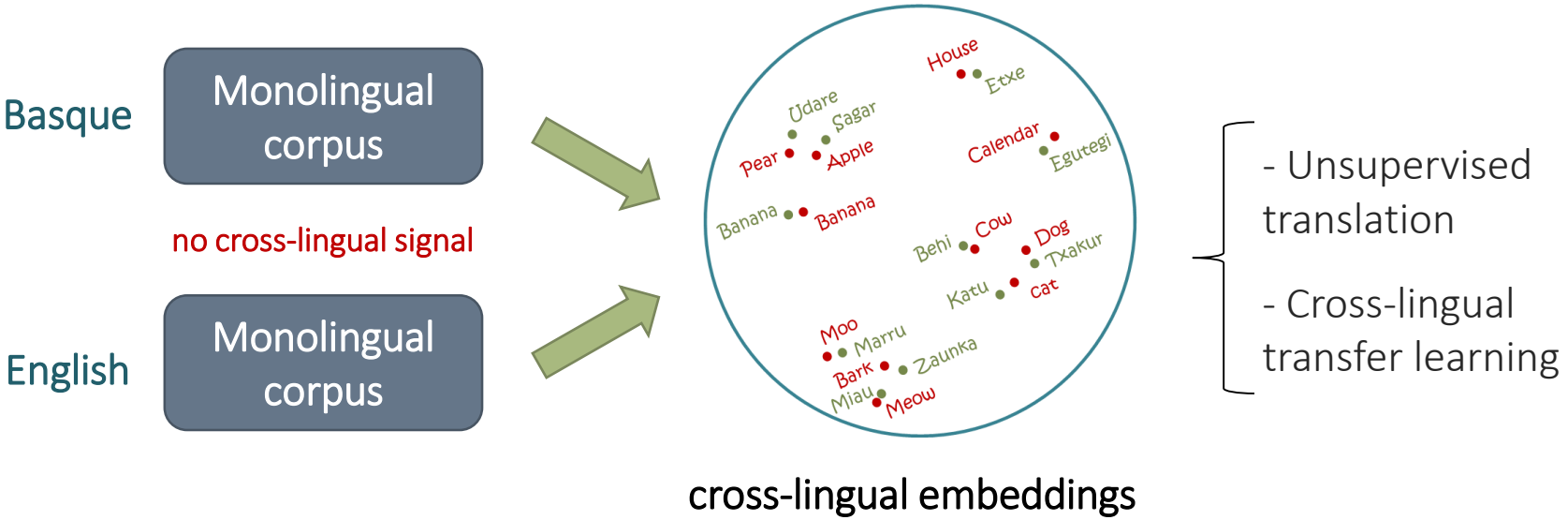
Introduction



Introduction



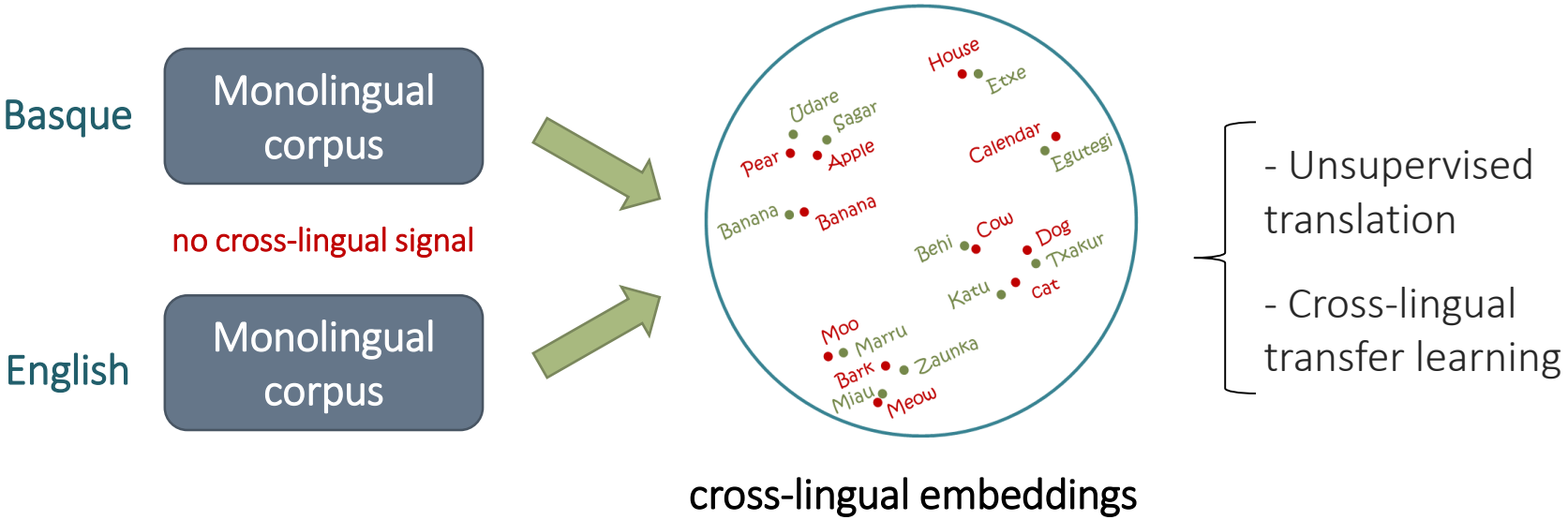
Introduction



Previous work

(Zhang et al., 2017; Conneau et al., 2018)

Introduction

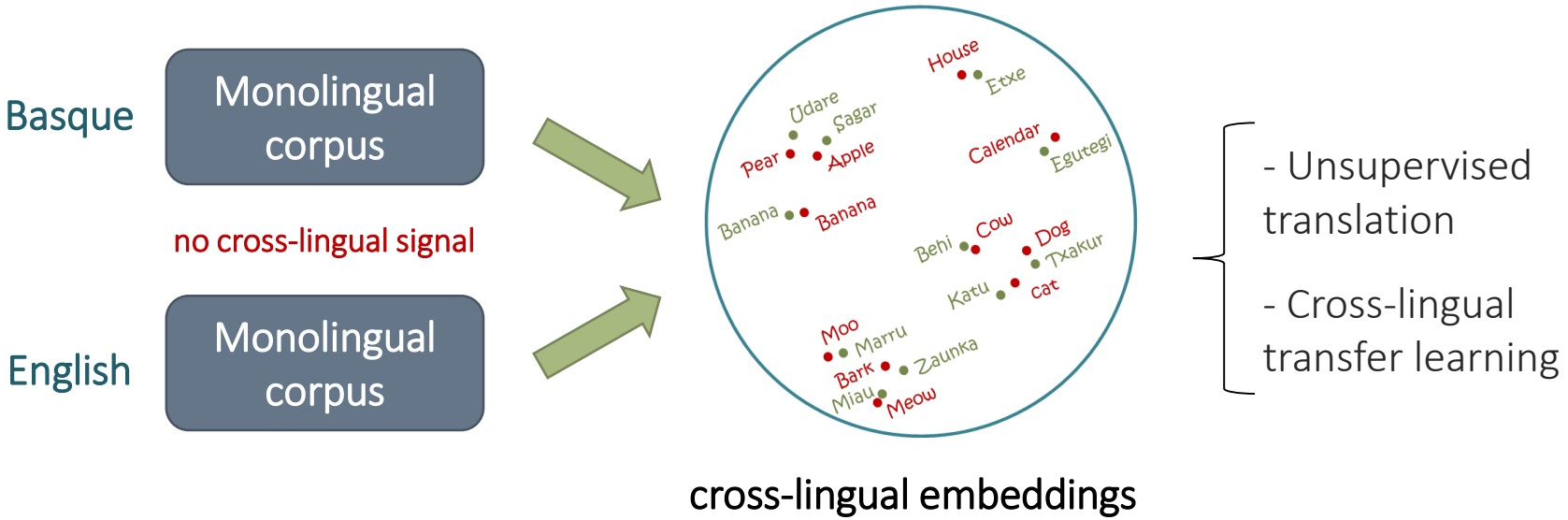


Previous work

(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning

Introduction



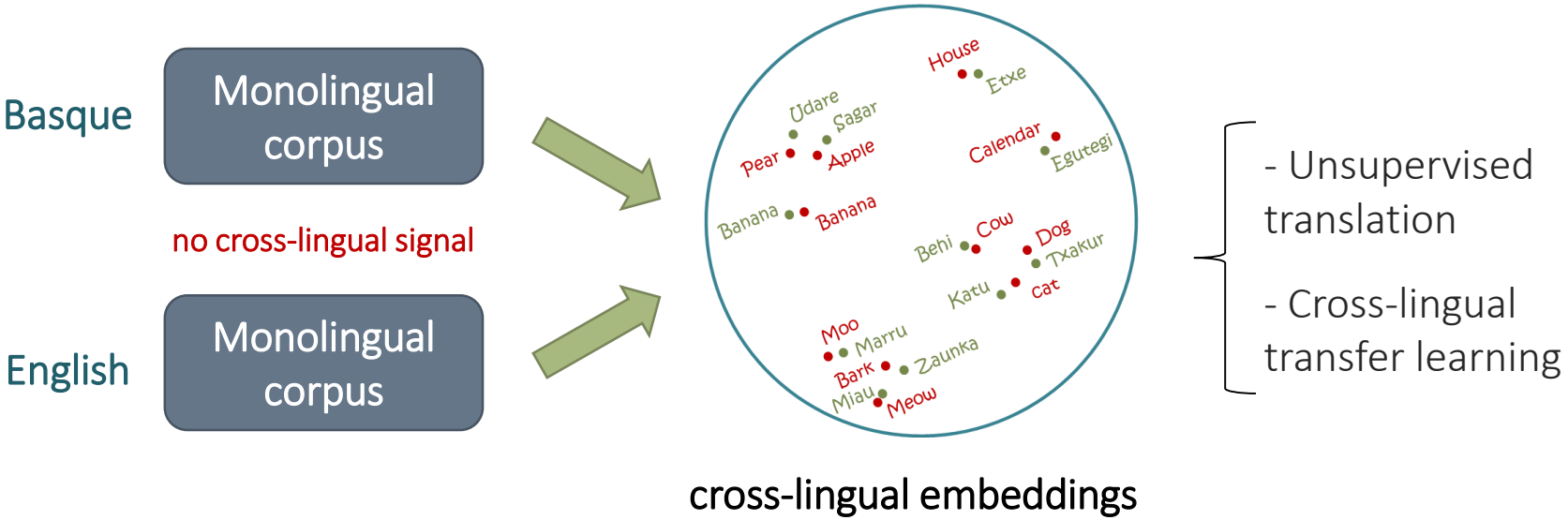
Previous work

(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning

- Very good results

Introduction

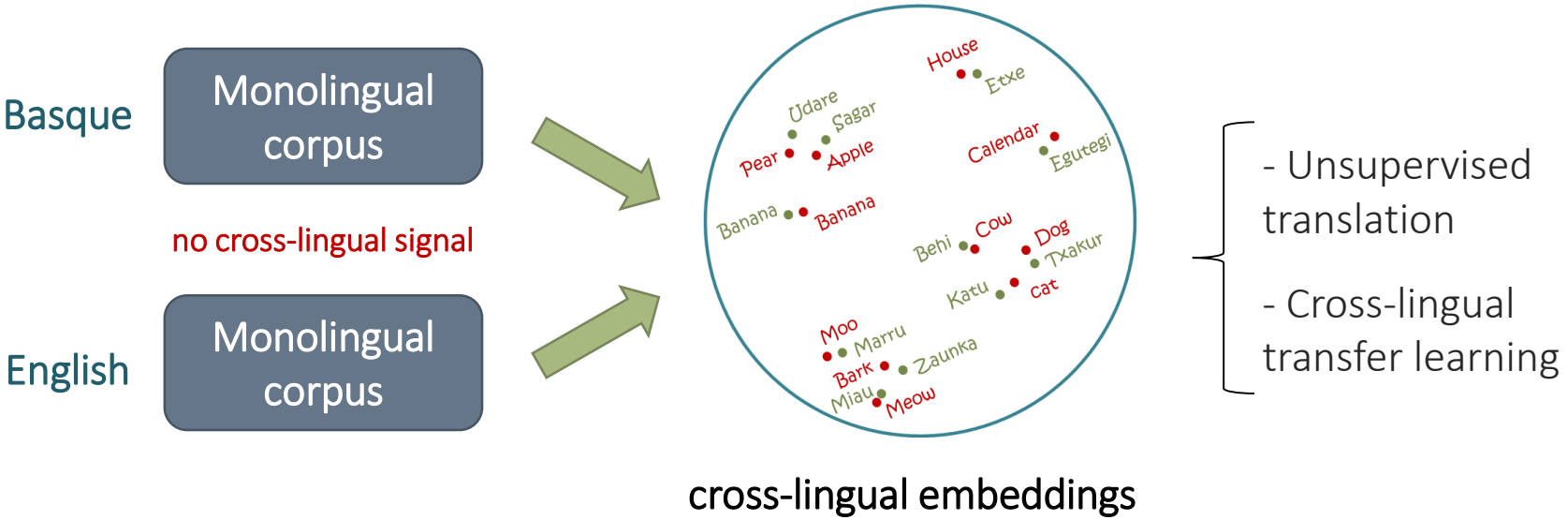


Previous work

(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning
- Very good results
- Tested in favorable conditions

Introduction

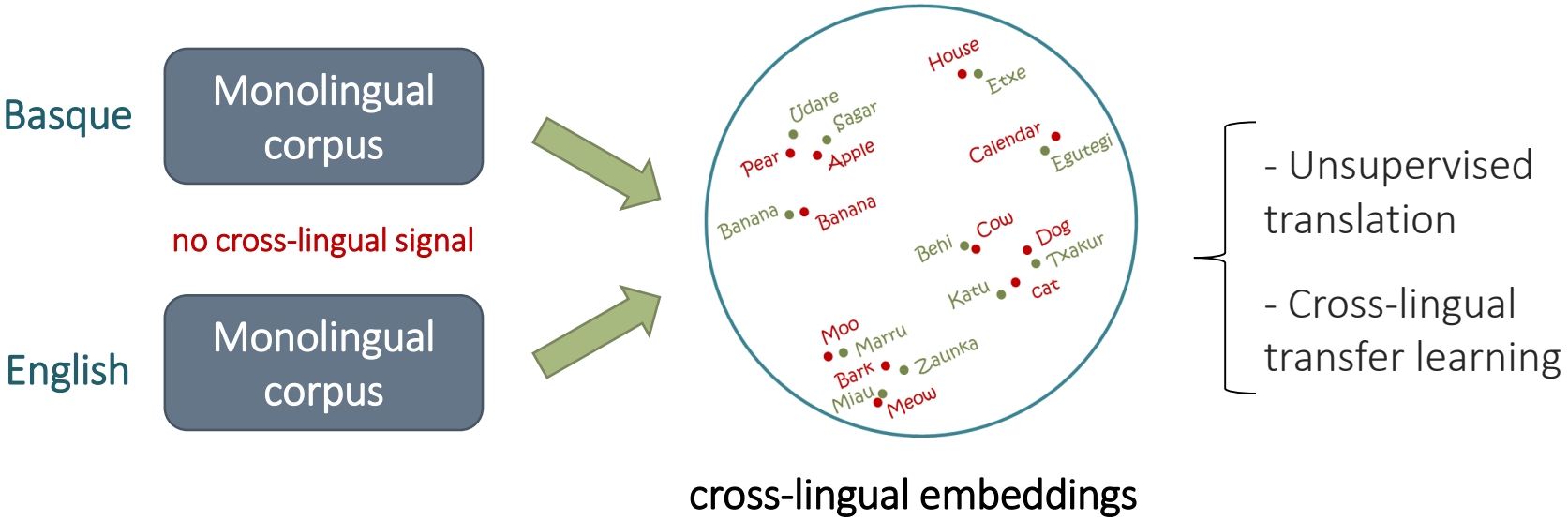


Previous work

(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning
- Very good results
- Tested in favorable conditions
  - Fail in more challenging datasets

Introduction



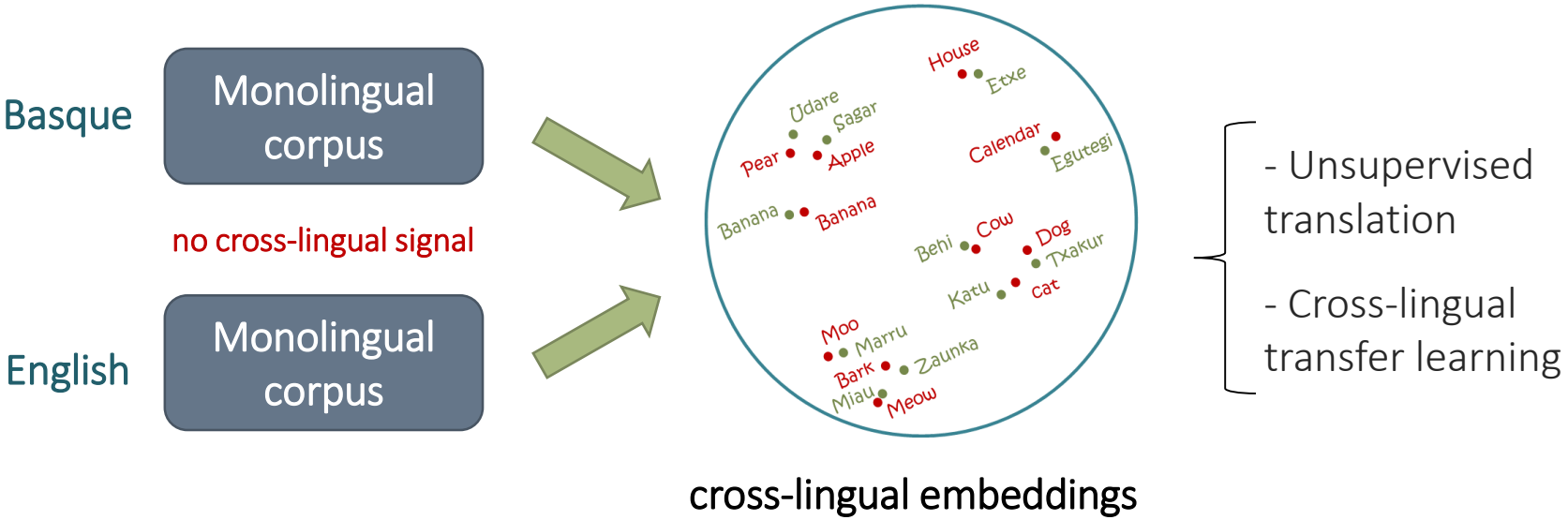
Previous work

(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning
- Very good results
- Tested in favorable conditions
  - Fail in more challenging datasets

This work

Introduction



Previous work

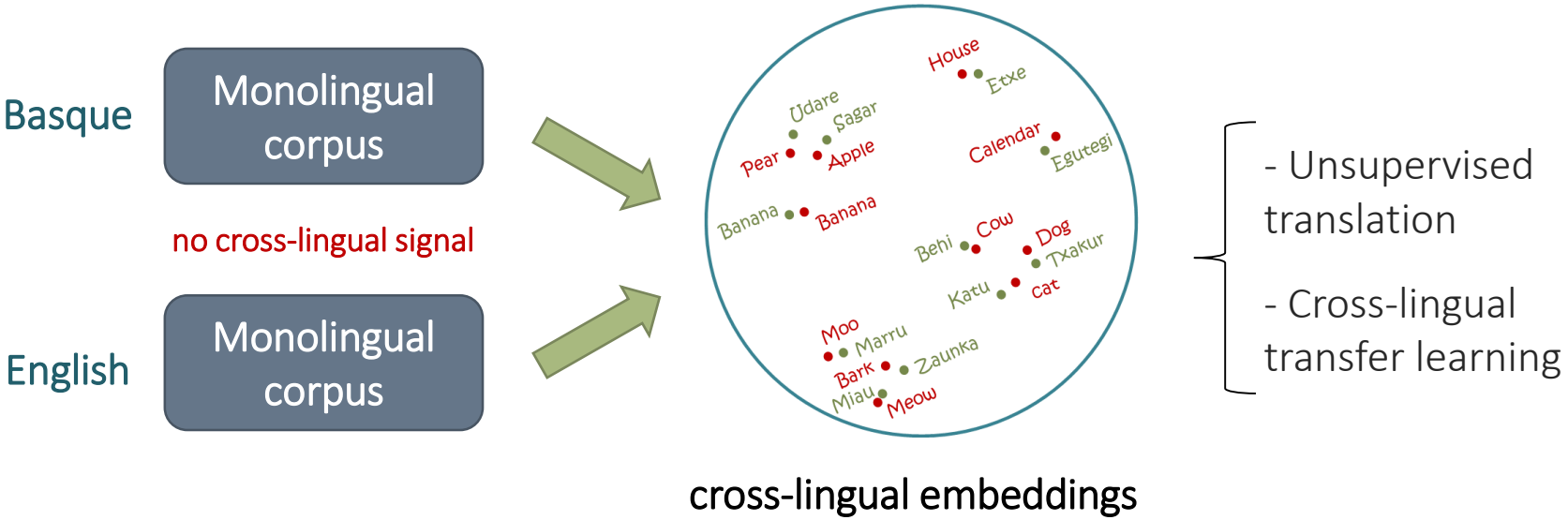
(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning
- Very good results
- Tested in favorable conditions
  - Fail in more challenging datasets

This work

- Self-learning

Introduction



Previous work

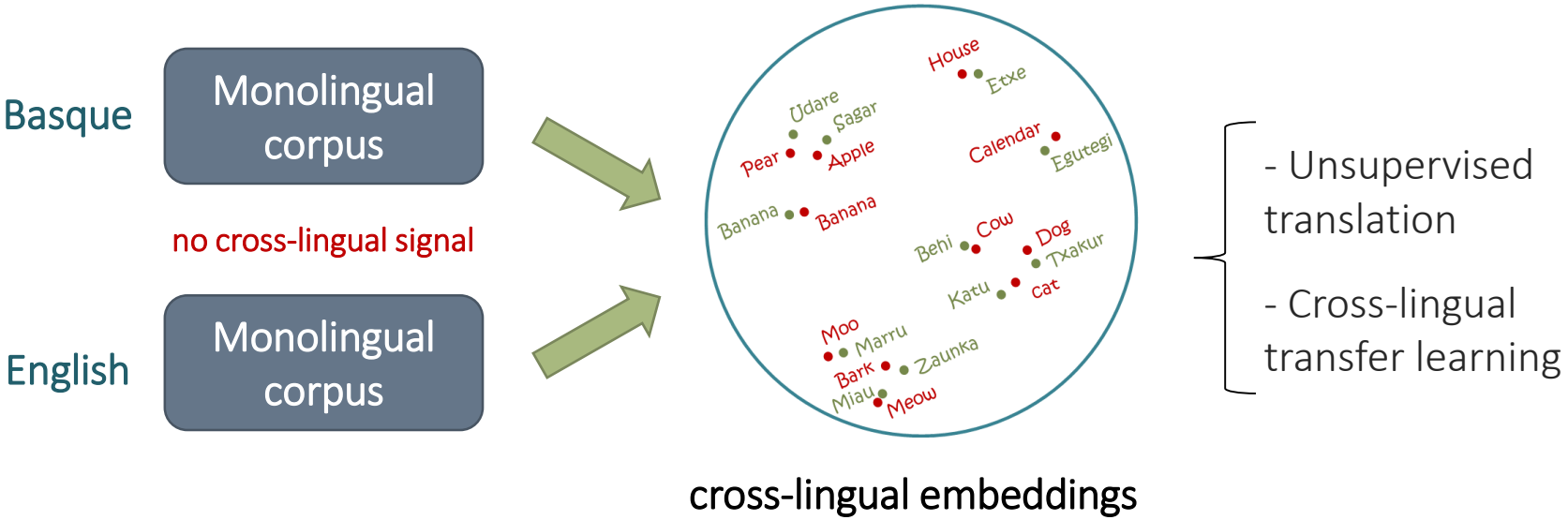
(Zhang et al., 2017; Conneau et al., 2018)

- Adversarial learning
- Very good results
- Tested in favorable conditions
  - Fail in more challenging datasets

This work

- Self-learning
- Even better results

Introduction



Previous work

(Zhang et al., 2017; Conneau et al., 2018)

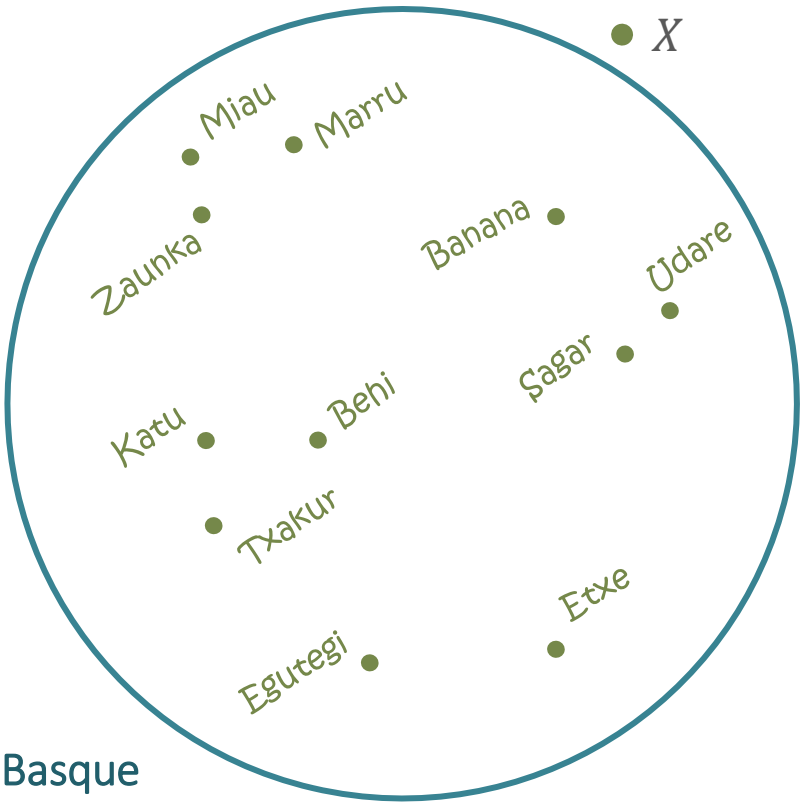
- Adversarial learning
- Very good results
- Tested in favorable conditions
  - Fail in more challenging datasets

This work

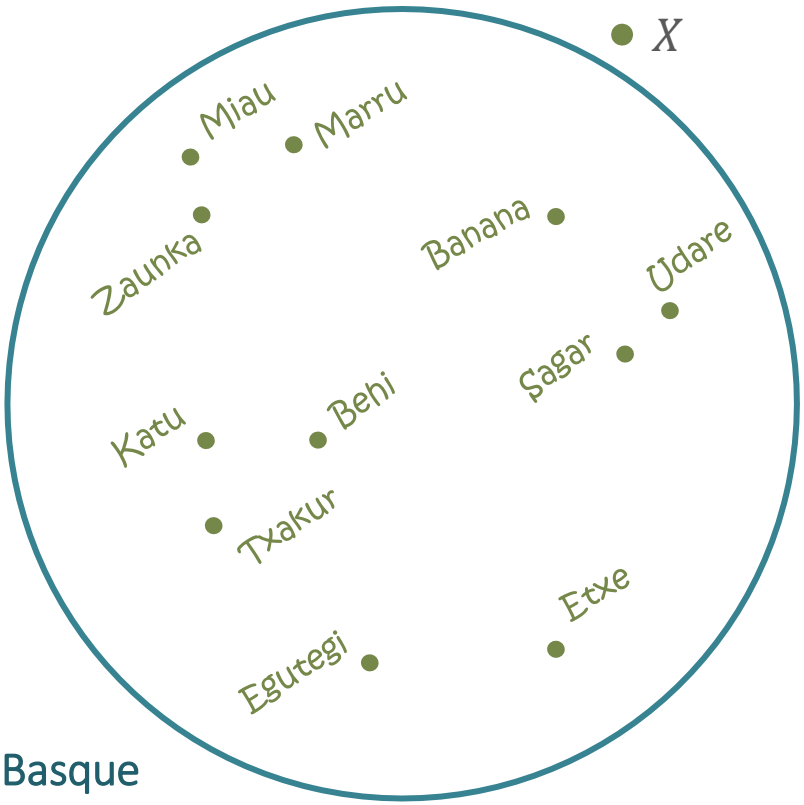
- Self-learning
- Even better results
- Works in challenging datasets

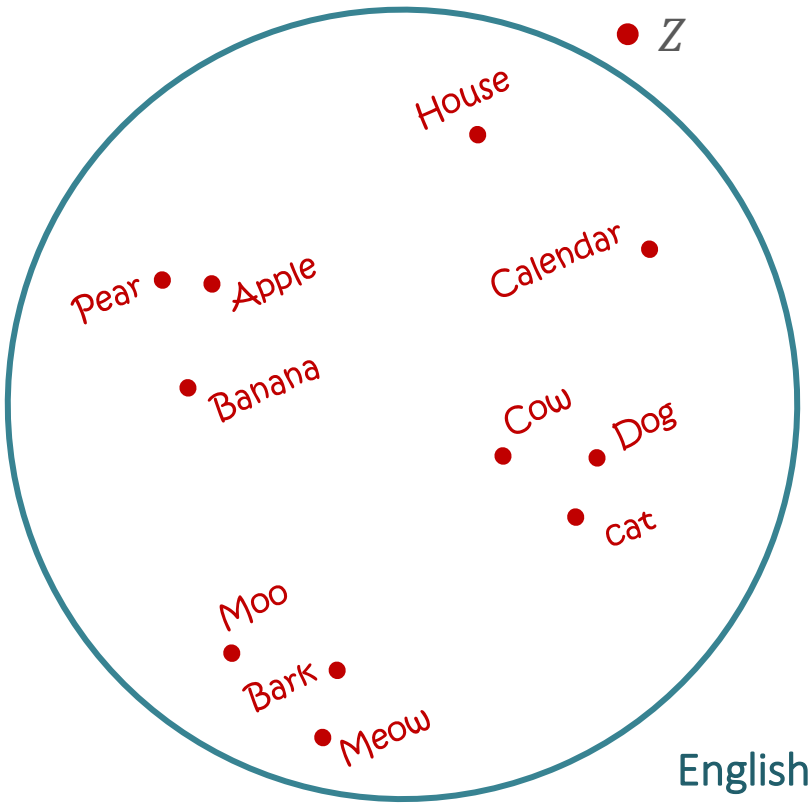
Cross-lingual embedding mappings

Cross-lingual embedding mappings

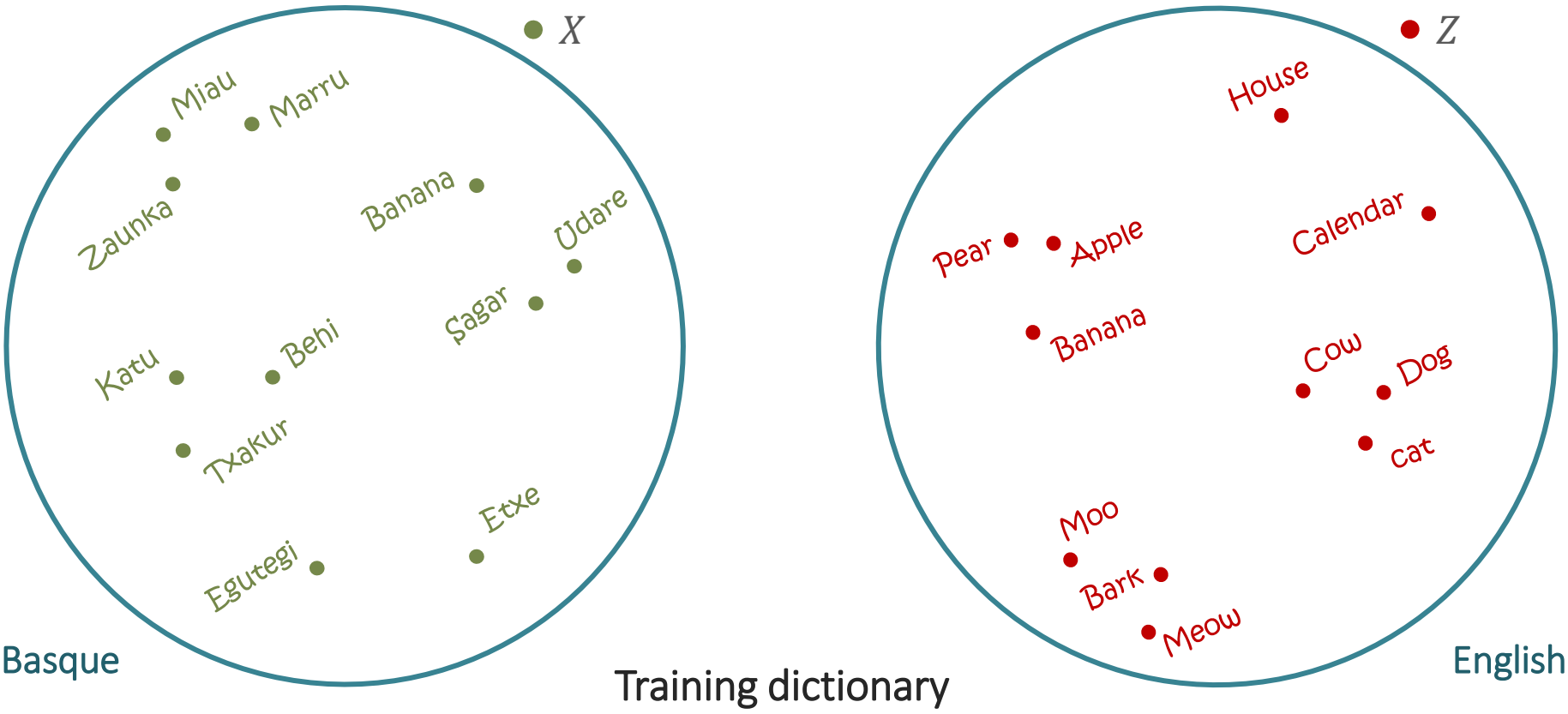


Cross-lingual embedding mappings



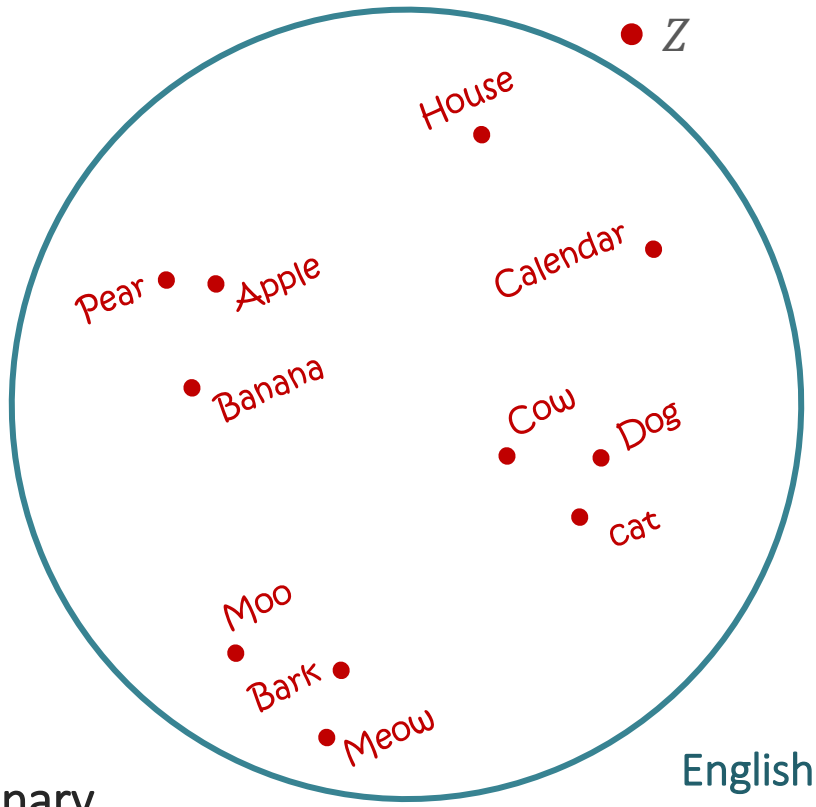


Cross-lingual embedding mappings



Cross-lingual embedding mappings

Basque

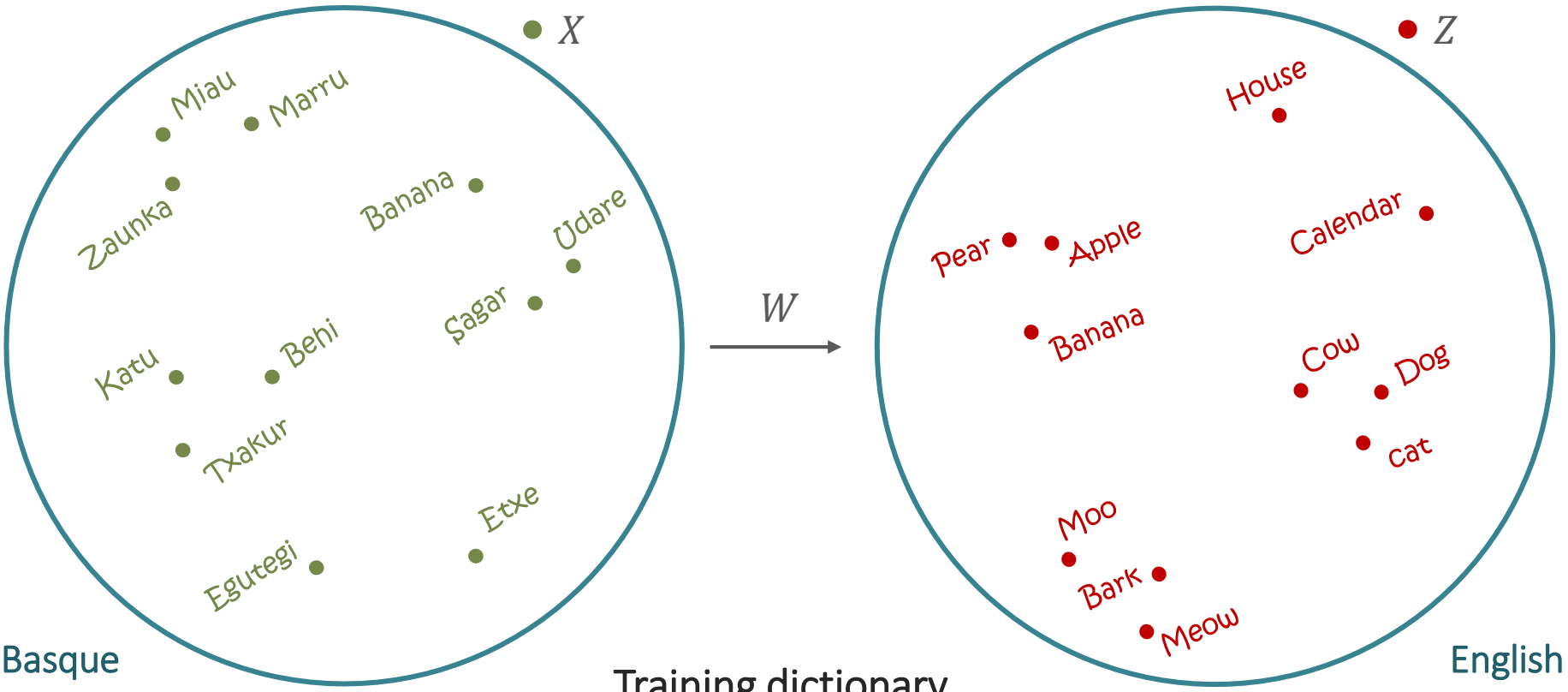


English

Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

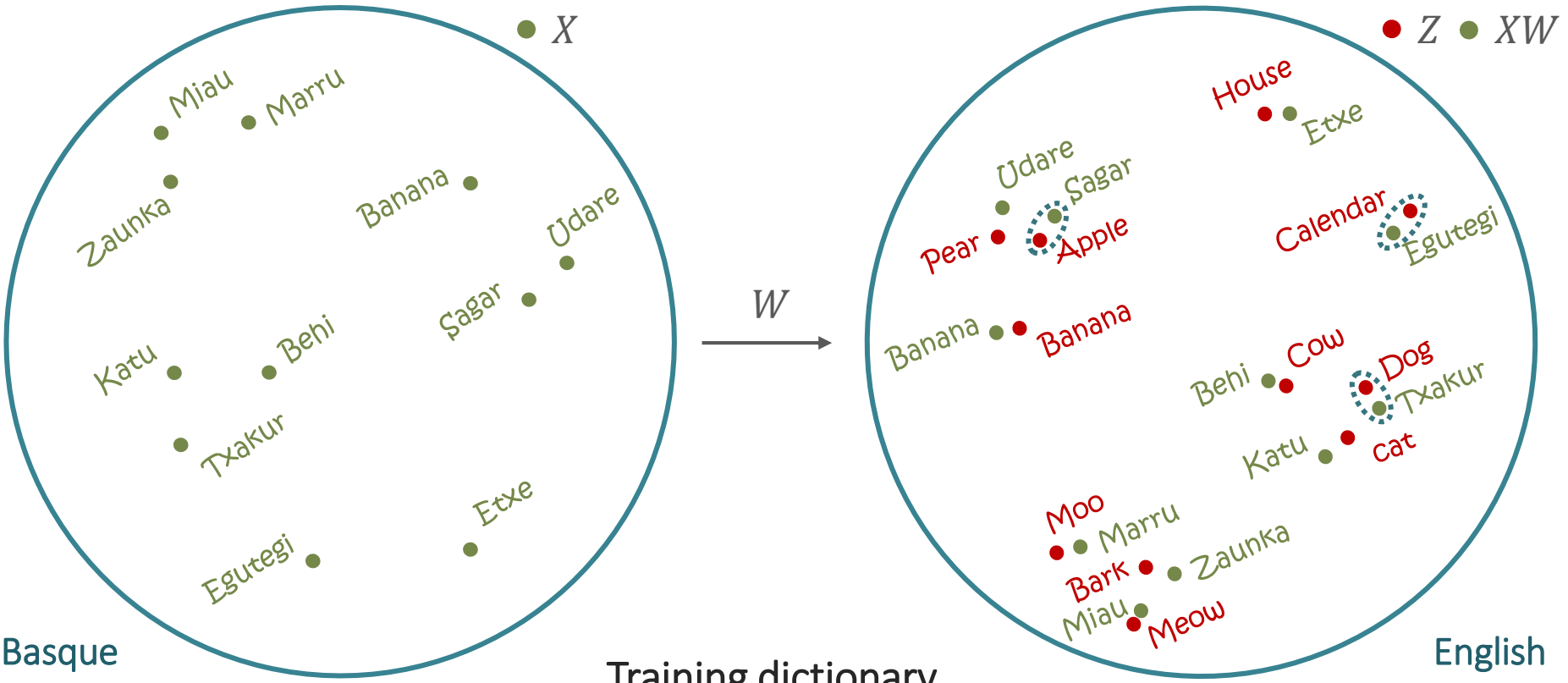
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

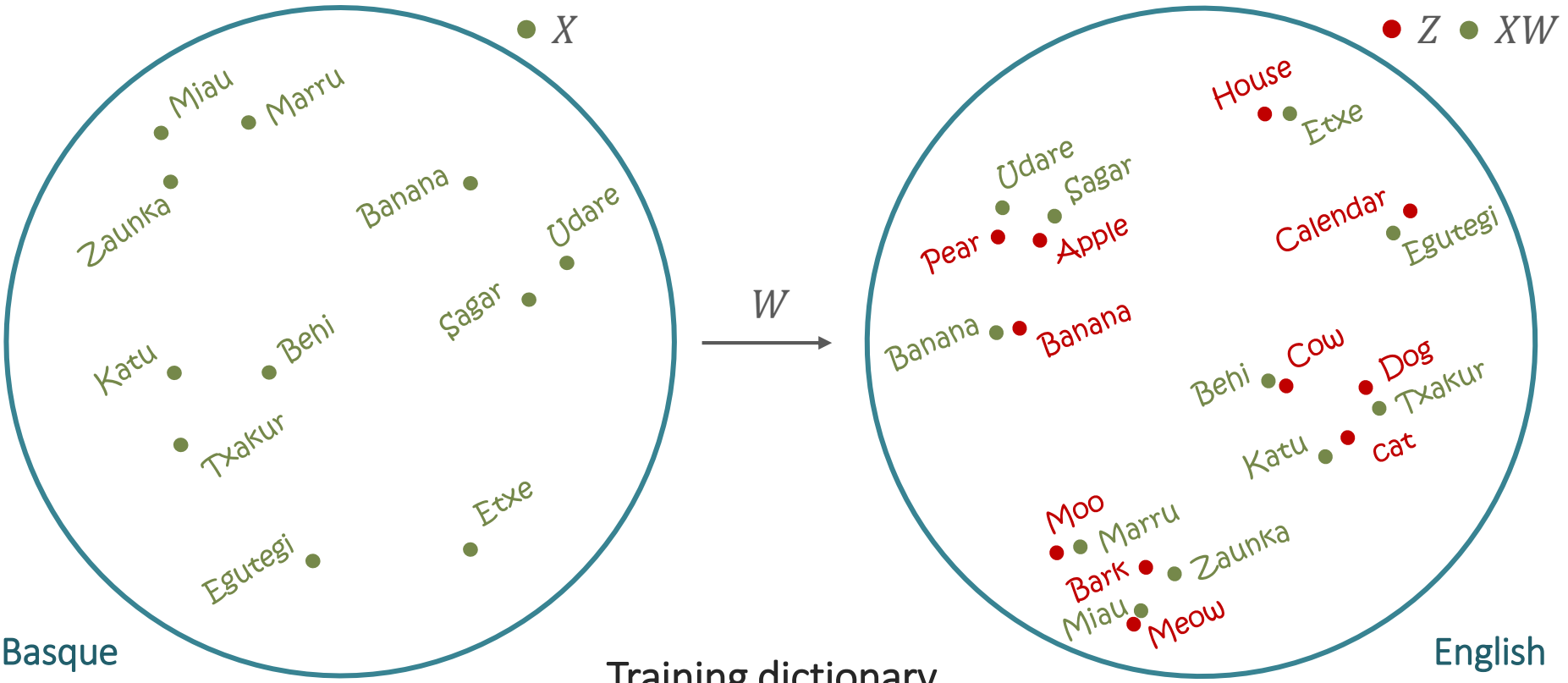
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

Cross-lingual embedding mappings



Training dictionary

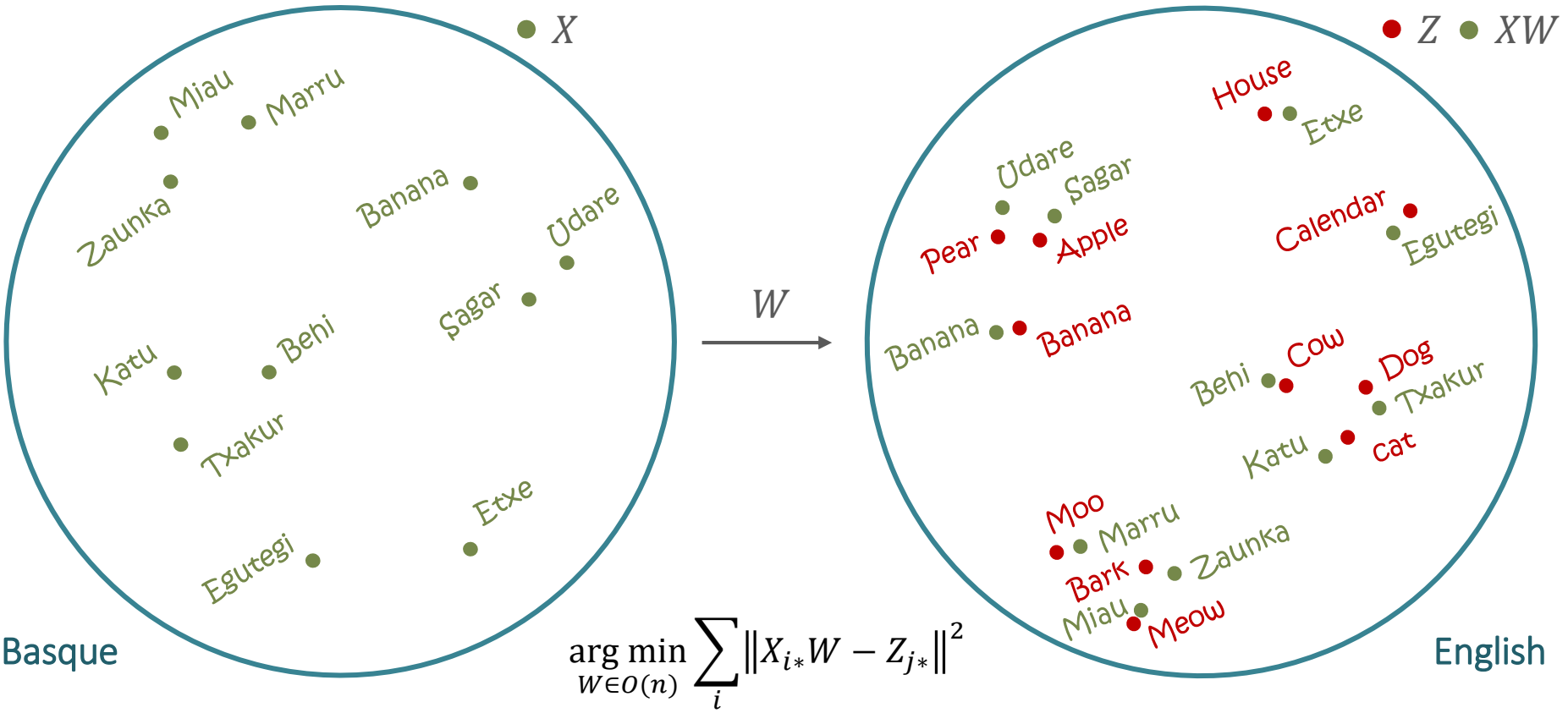
$$\begin{matrix}\text{Txakur}\\ \text{Sagar}\\ \vdots\\ \text{Egutegi}\end{matrix}\begin{bmatrix}X_{1,*}\\ X_{2,*}\\ \vdots\\ X_{n,*}\end{bmatrix} \quad \Big| \quad \begin{bmatrix}Z_{1,*}\\ Z_{2,*}\\ \vdots\\ Z_{n,*}\end{bmatrix}\begin{matrix}\text{Dog}\\ \text{Apple}\\ \vdots\\ \text{Calendar}\end{matrix}$$

Cross-lingual embedding mappings

Training dictionary

$$\begin{matrix}\text{Txakur}\\ \text{Sagar}\\ \vdots\\ \text{Egutegi}\end{matrix}\begin{bmatrix}X_{1,*}\\ X_{2,*}\\ \vdots\\ X_{n,*}\end{bmatrix}[W] \approx \begin{bmatrix}Z_{1,*}\\ Z_{2,*}\\ \vdots\\ Z_{n,*}\end{bmatrix}\begin{matrix}\text{Dog}\\ \text{Apple}\\ \vdots\\ \text{Calendar}\end{matrix}$$

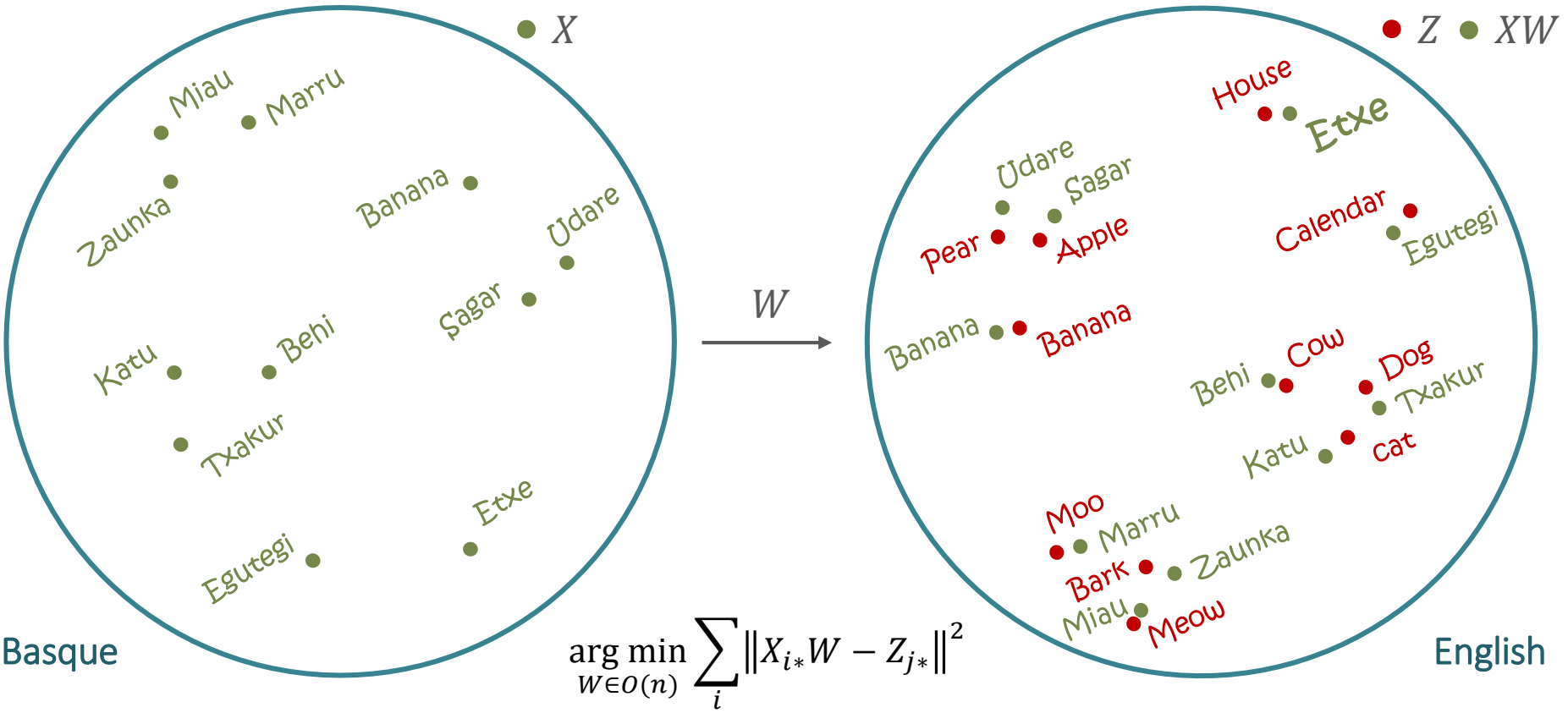
Cross-lingual embedding mappings



$$\arg\min_{W \in O(n)} \sum_i \|X_{i*}W - Z_{j*}\|^2$$

$$\begin{matrix} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{matrix} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} [W] \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{matrix} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{matrix}$$

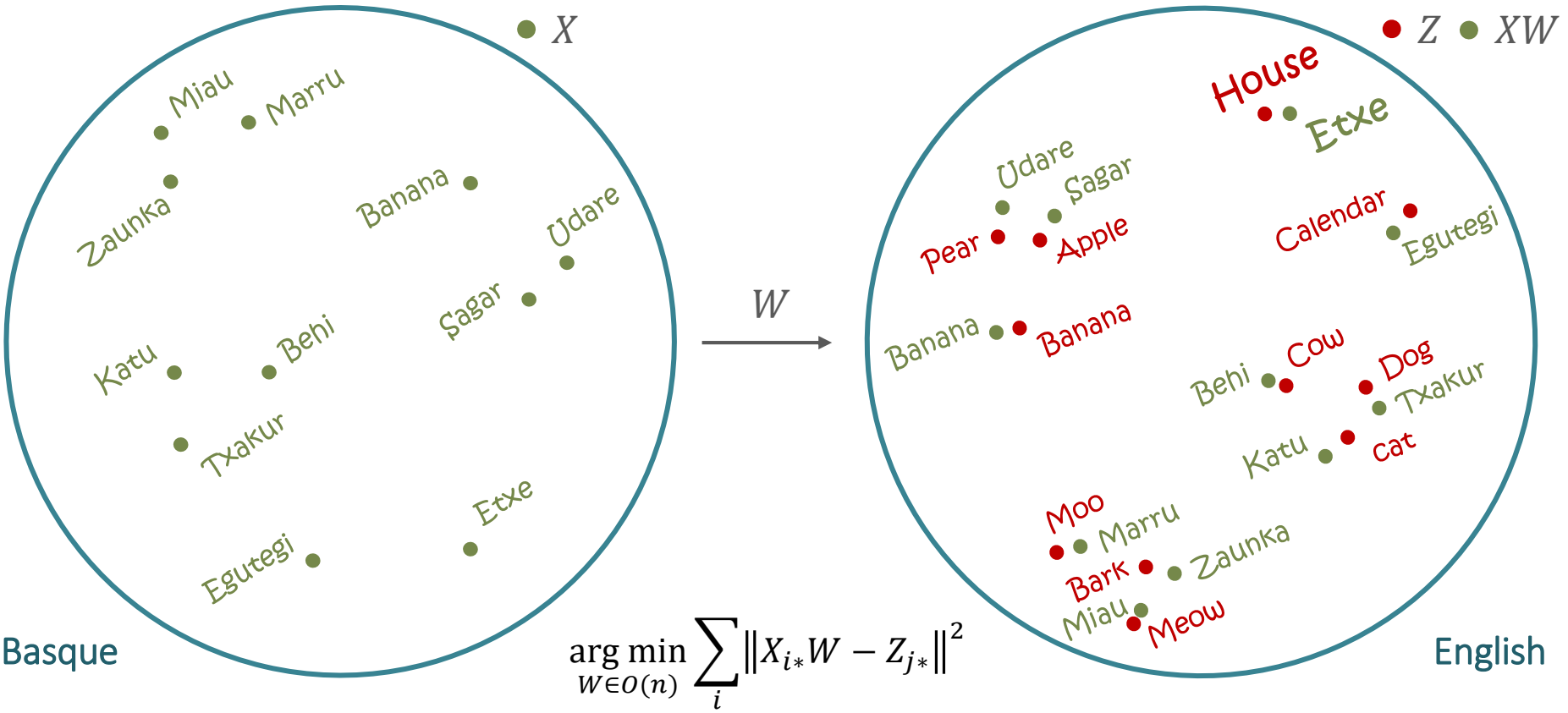
Cross-lingual embedding mappings



$$\arg\min_{W \in O(n)} \sum_i \|X_{i*}W - Z_{j*}\|^2$$

$$\begin{matrix} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{matrix} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} [W] \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{matrix} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{matrix}$$

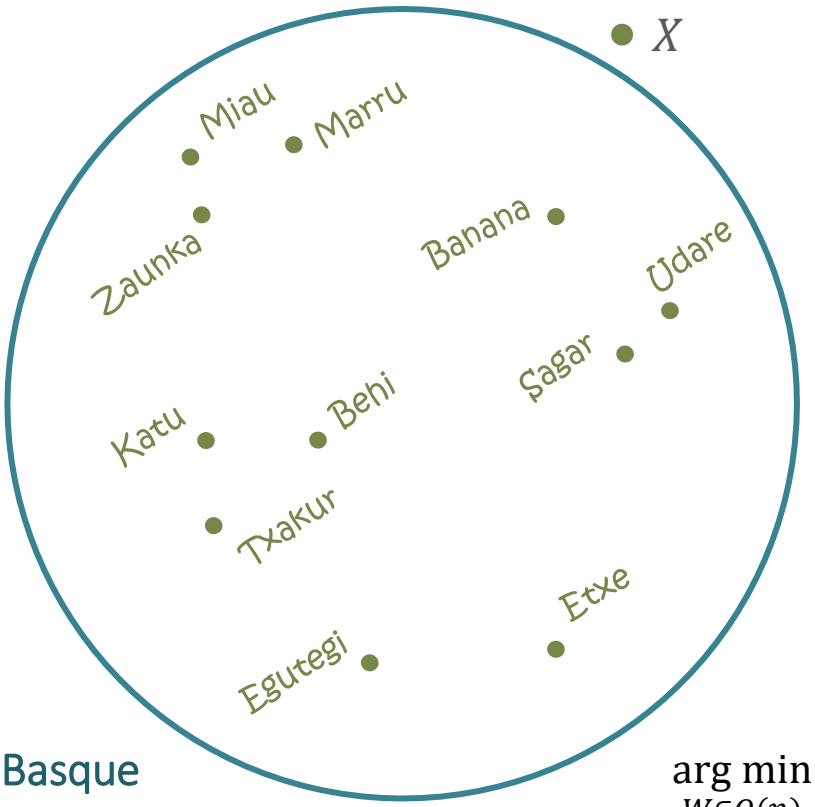
Cross-lingual embedding mappings



$$\arg\min_{W \in O(n)} \sum_i \|X_{i*}W - Z_{j*}\|^2$$

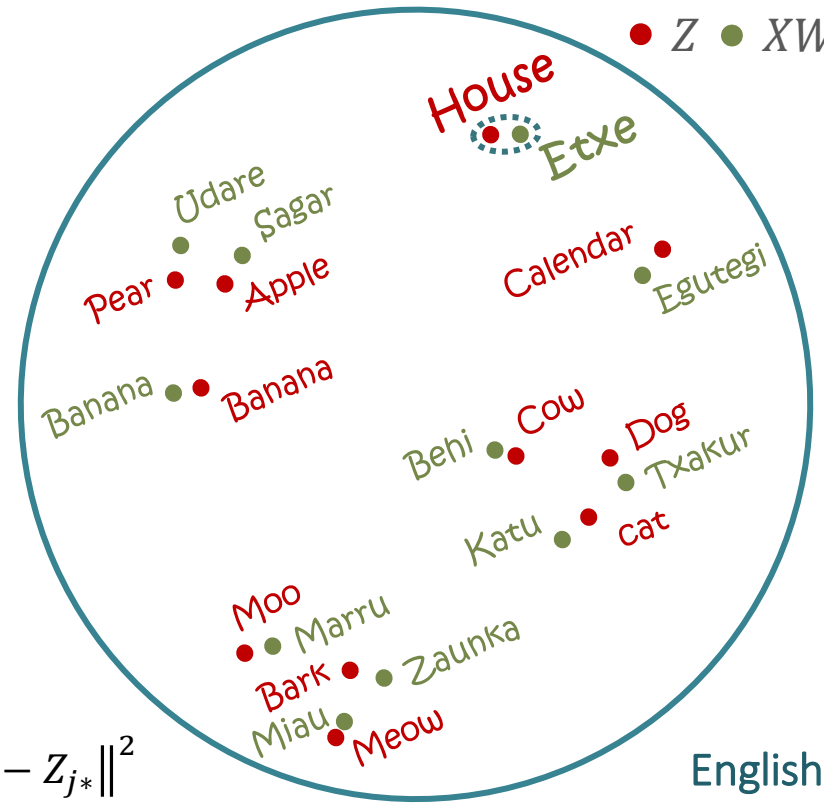
$$\begin{matrix} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{matrix} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} [W] \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{matrix} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{matrix}$$

Cross-lingual embedding mappings



Basque

W

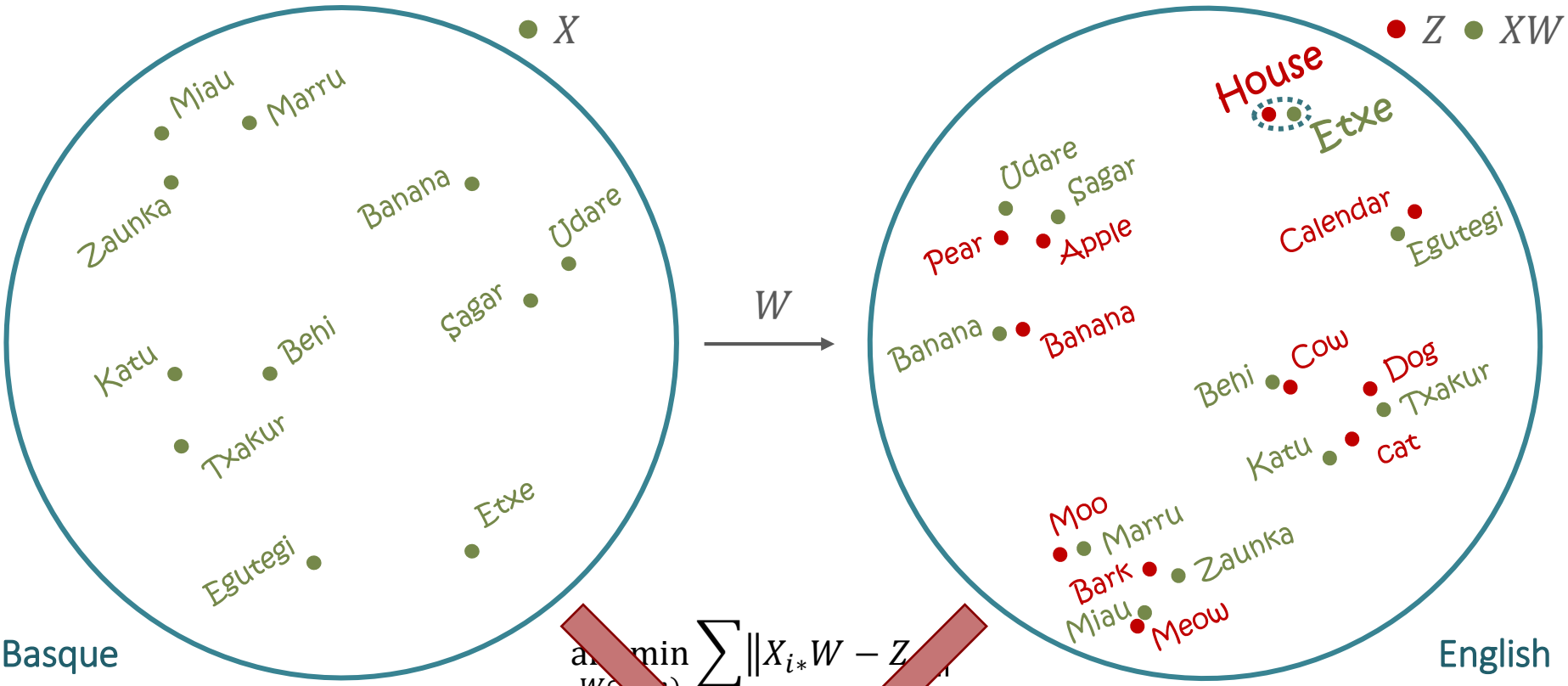


English

$$\arg\min_{W \in O(n)} \sum_i \|X_{i*}W - Z_{j*}\|^2$$

$$\begin{matrix} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{matrix} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} [W] \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{matrix} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{matrix}$$

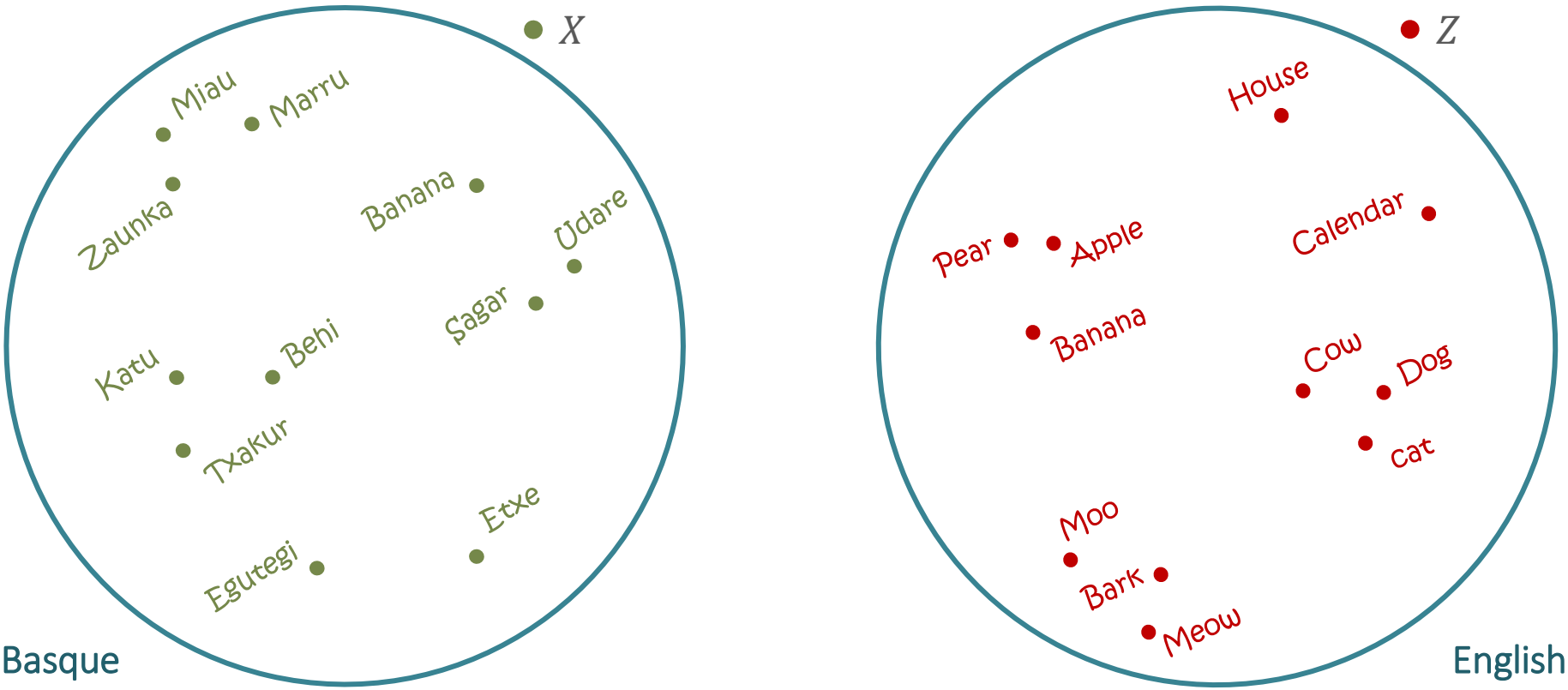
Cross-lingual embedding mappings



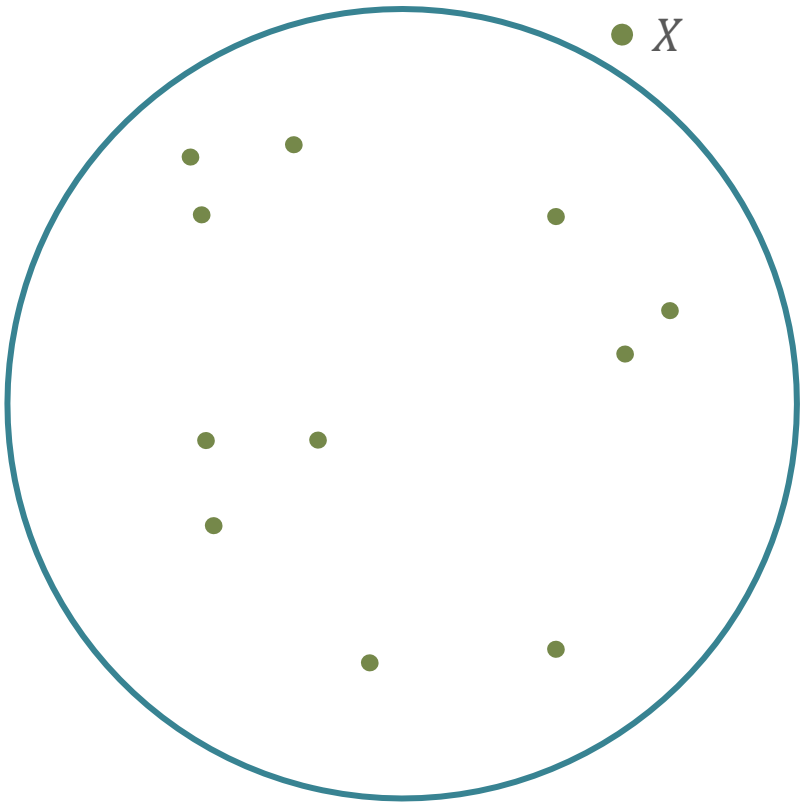
$$\underset{W}{\text{a}}\min \sum_i \|X_{i*}W - Z$$

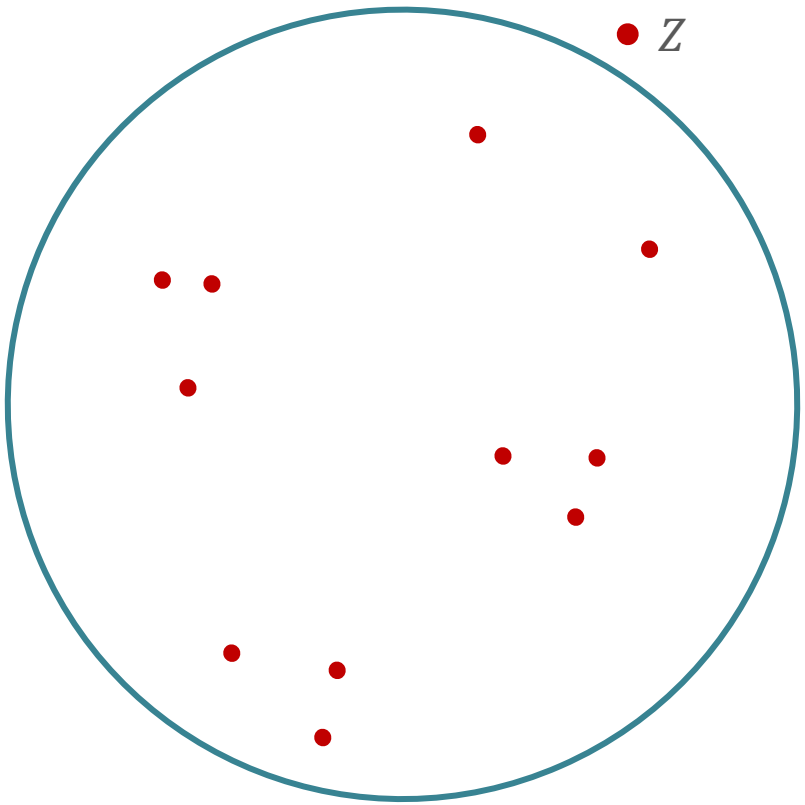
$$\begin{matrix}\text{Txakur}\\ \text{Sagar}\\ \vdots\\ \text{Egutegi}\end{matrix}\begin{bmatrix}X_{1,*}\\ X_{2,*}\\ \vdots\\ X_{\cdot,*}\end{bmatrix}\begin{bmatrix}\cdot\\ \cdot\end{bmatrix}\approx\begin{bmatrix}Z_{1,*}\\ Z_{2,*}\\ \vdots\\ Z_{\cdot,*}\end{bmatrix}\begin{matrix}\text{Dog}\\ \text{Apple}\\ \vdots\\ \text{Calendar}\end{matrix}$$

Cross-lingual embedding mappings

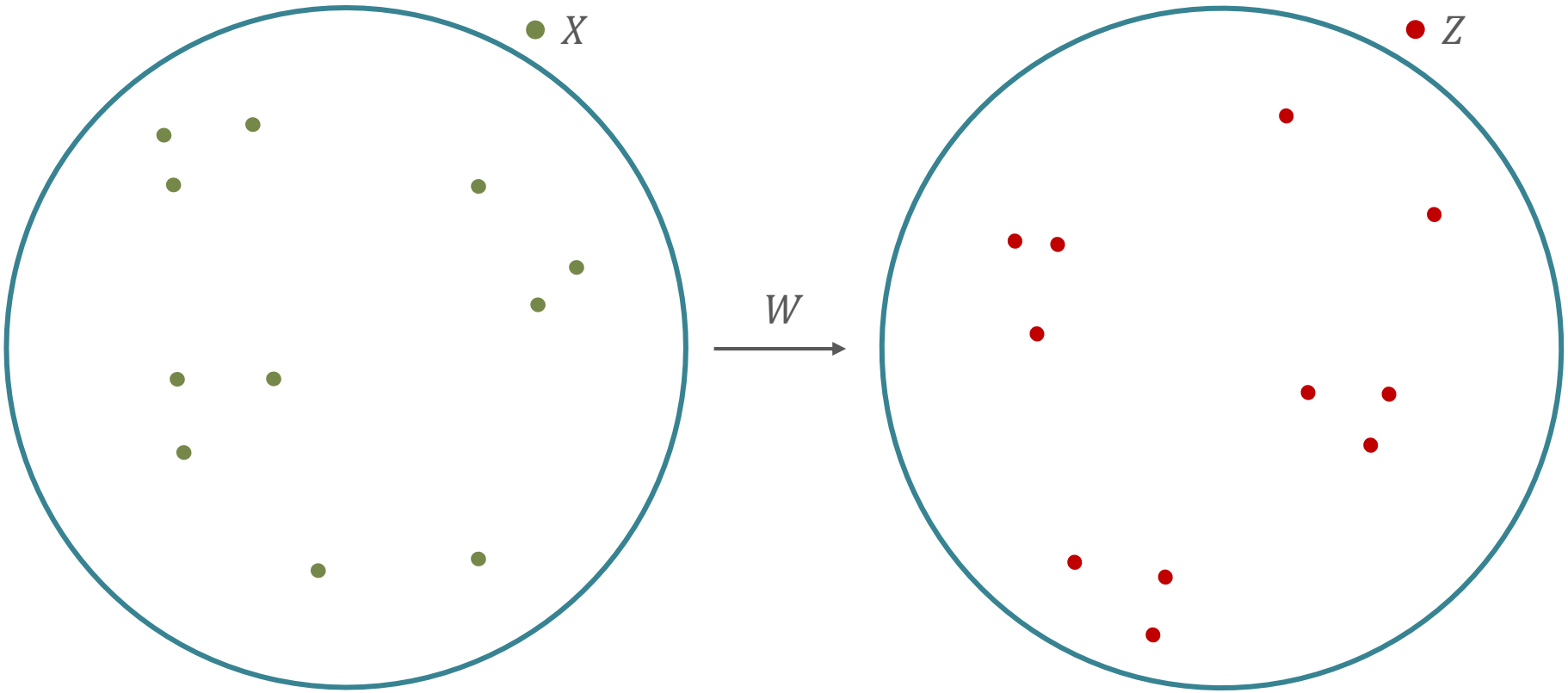


Cross-lingual embedding mappings

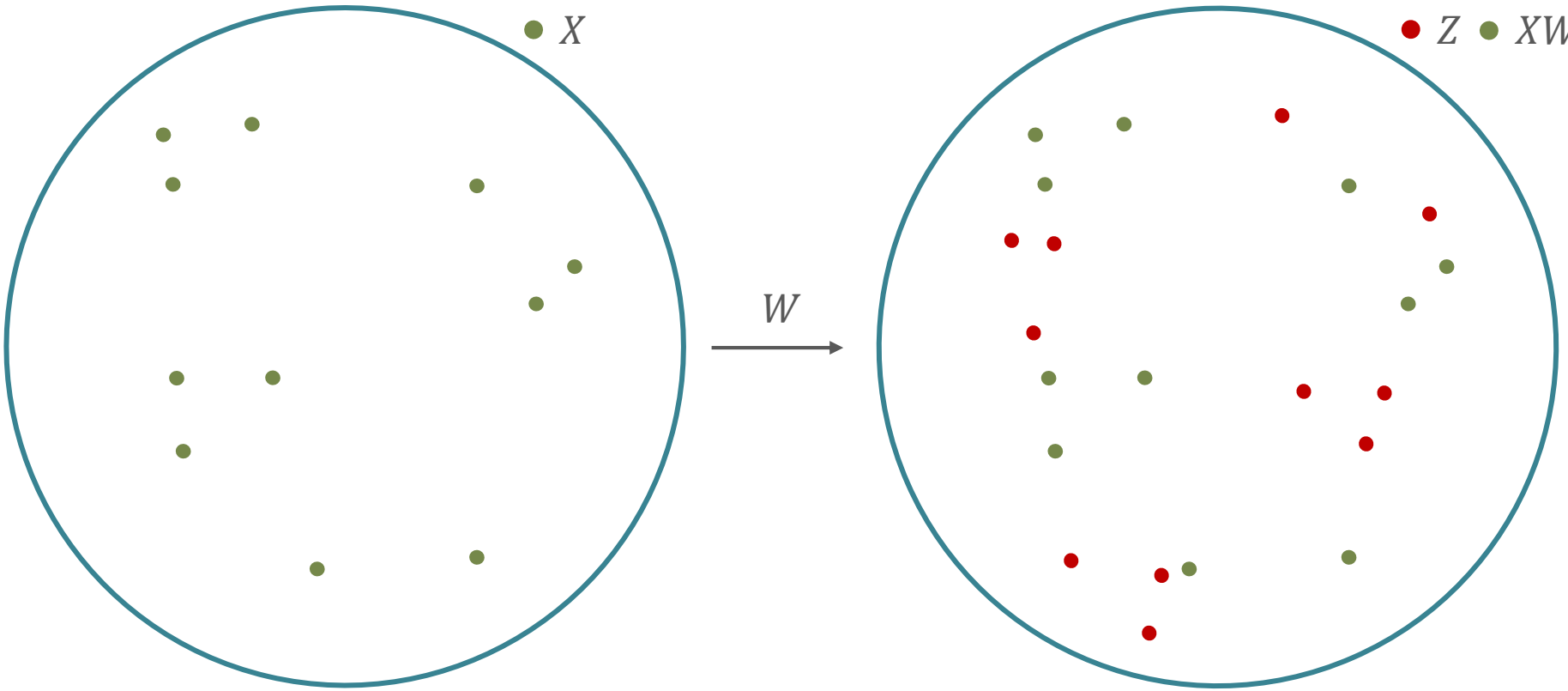




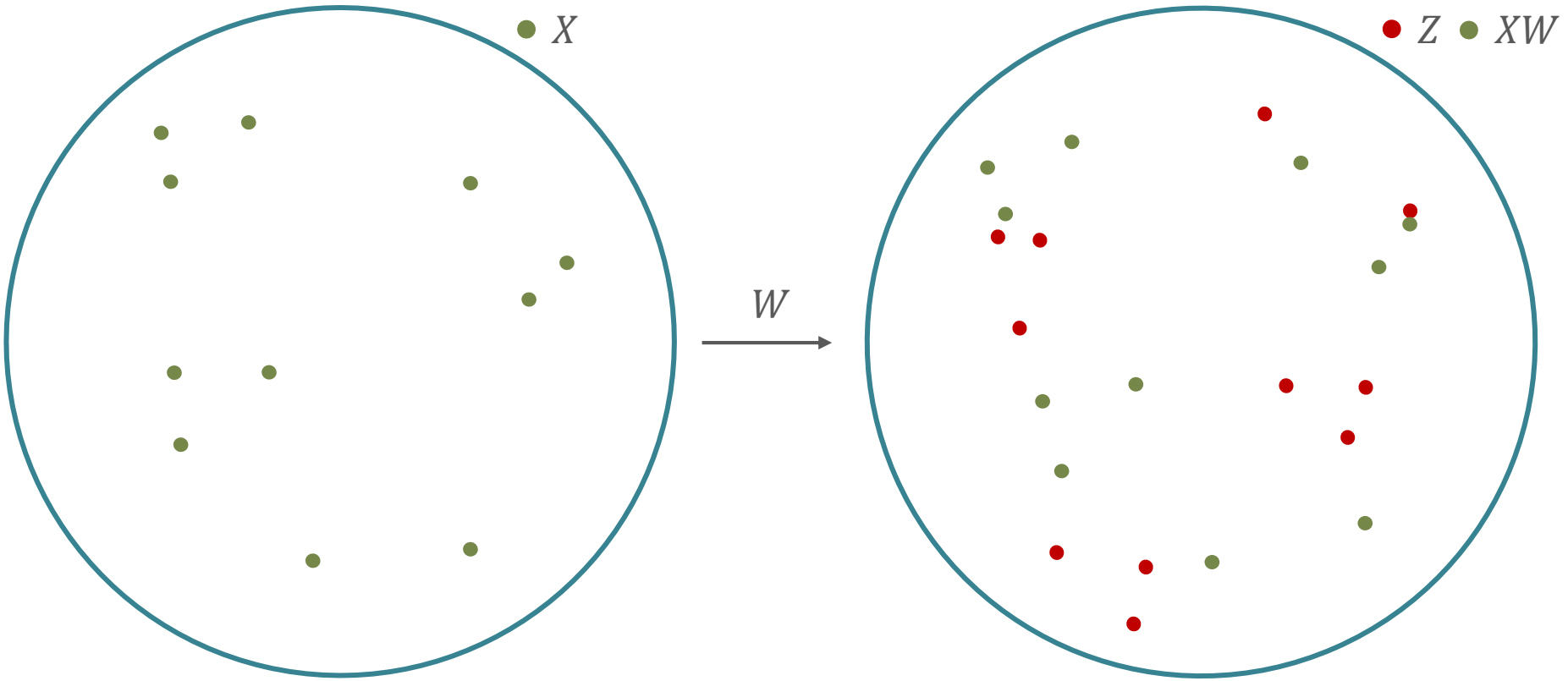
Cross-lingual embedding mappings



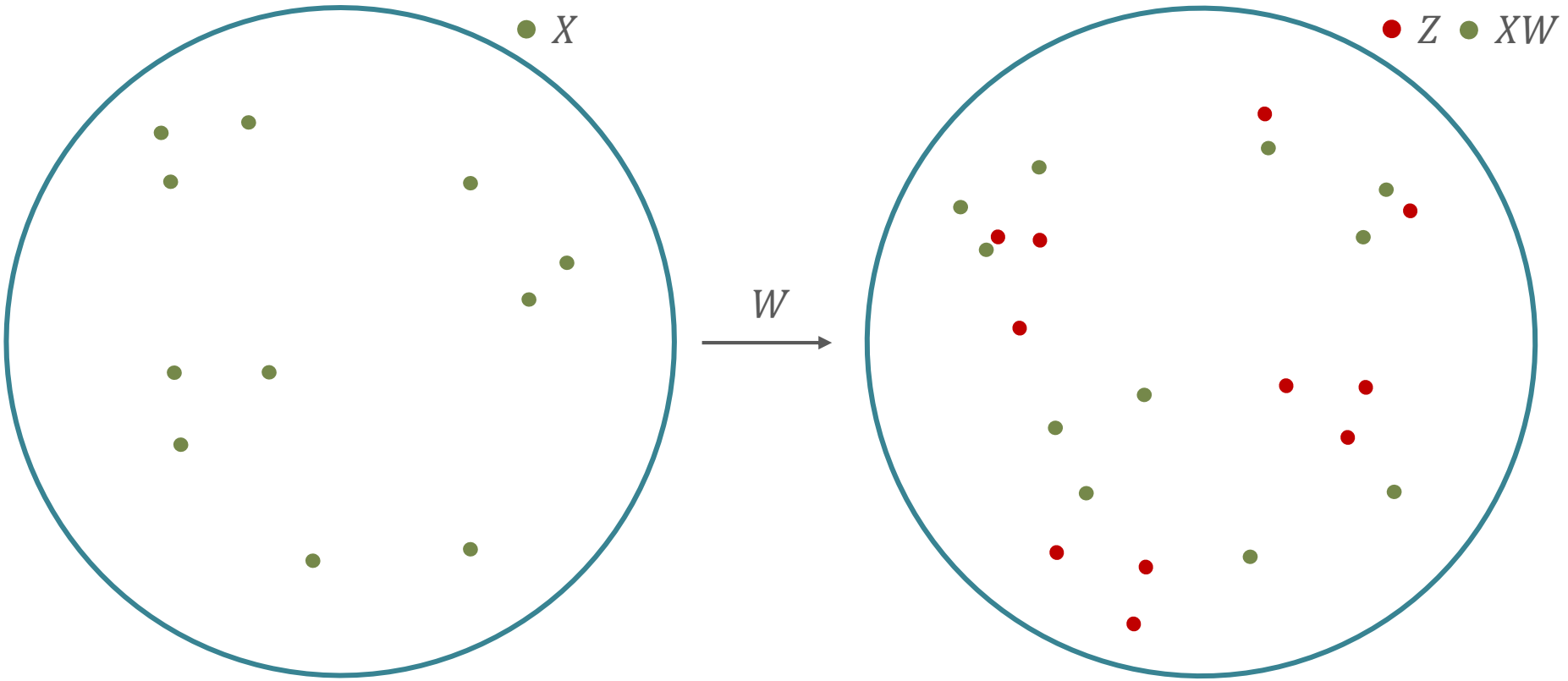
Cross-lingual embedding mappings



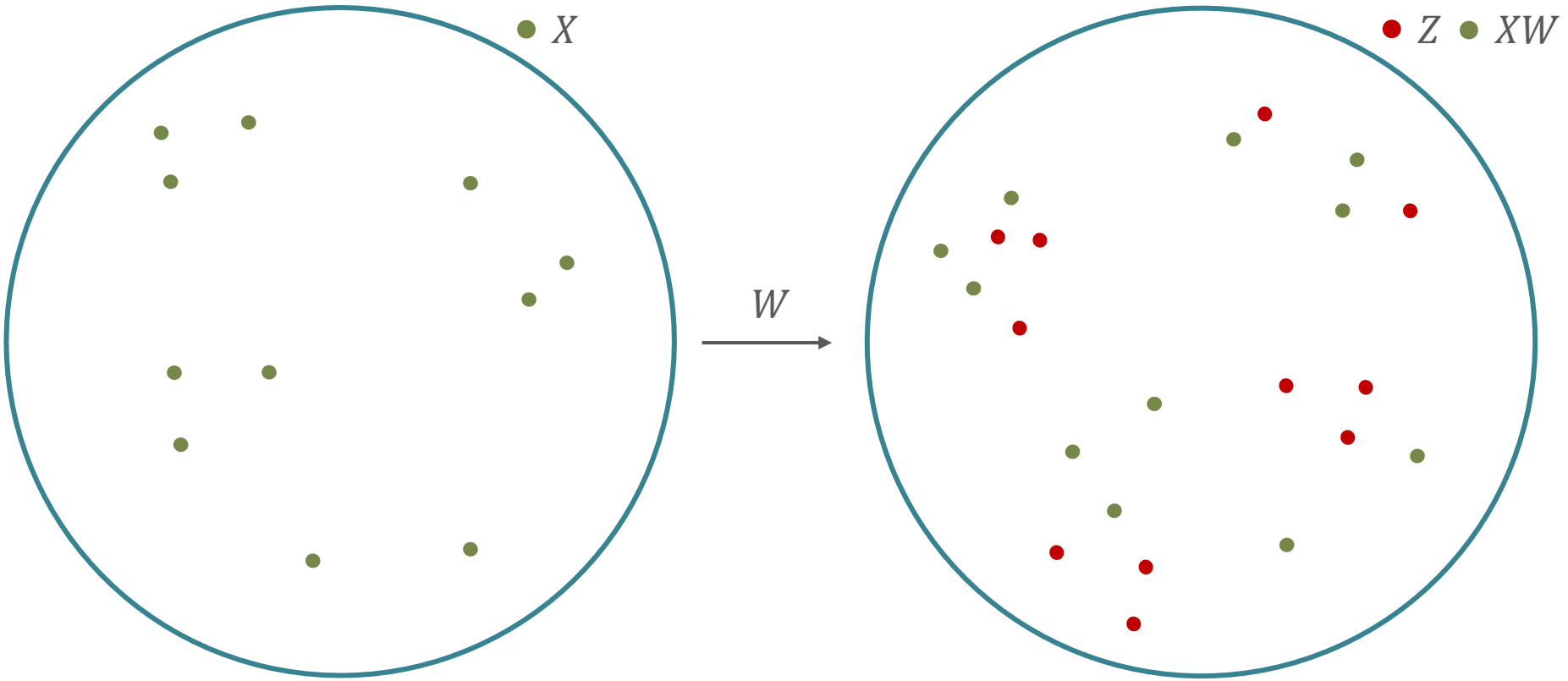
Cross-lingual embedding mappings



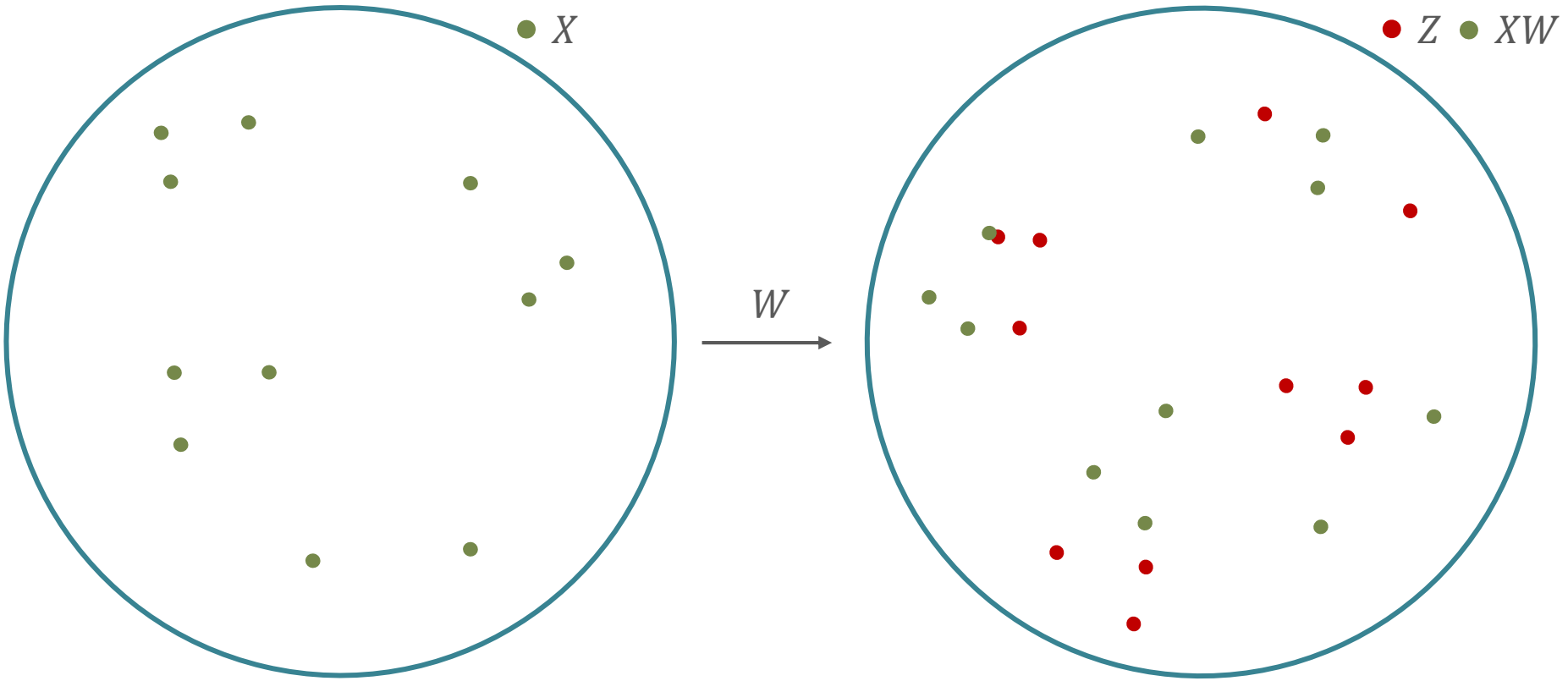
Cross-lingual embedding mappings



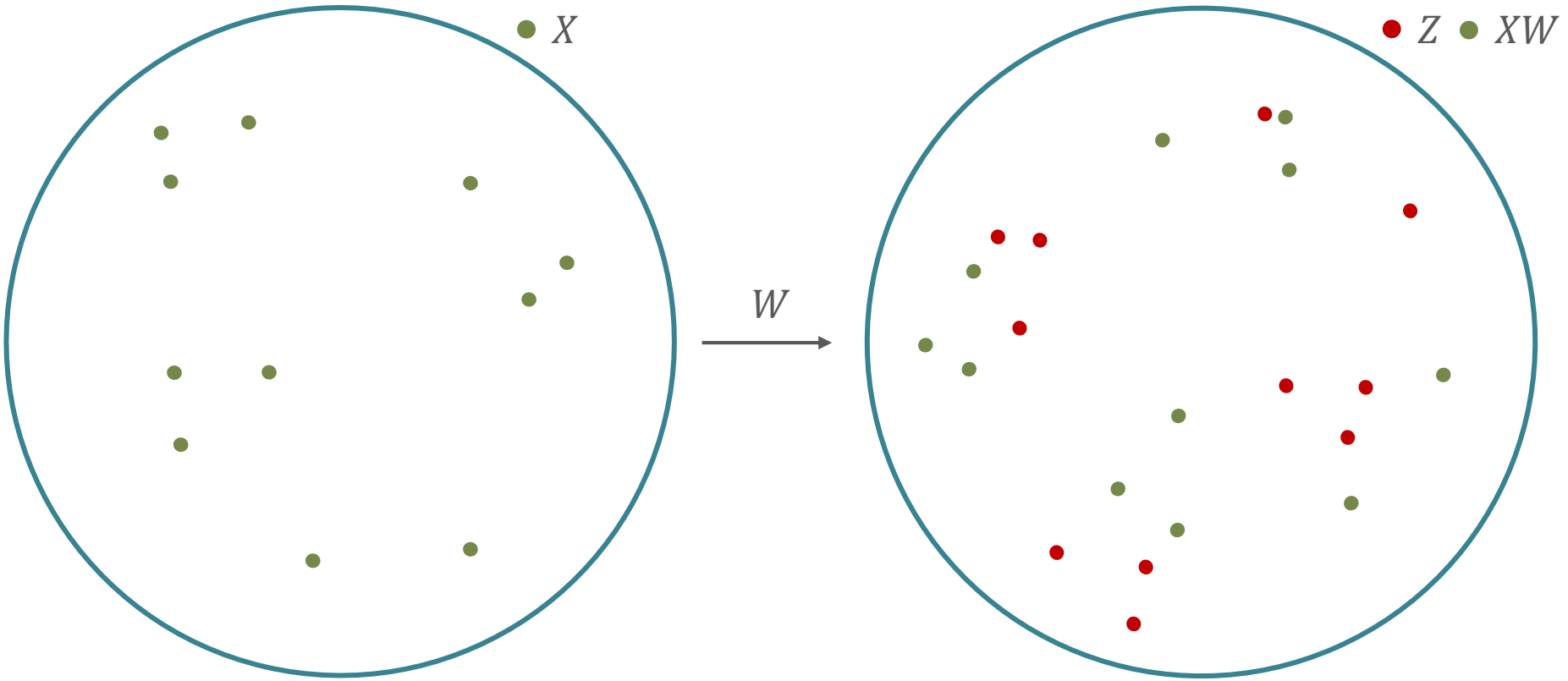
Cross-lingual embedding mappings



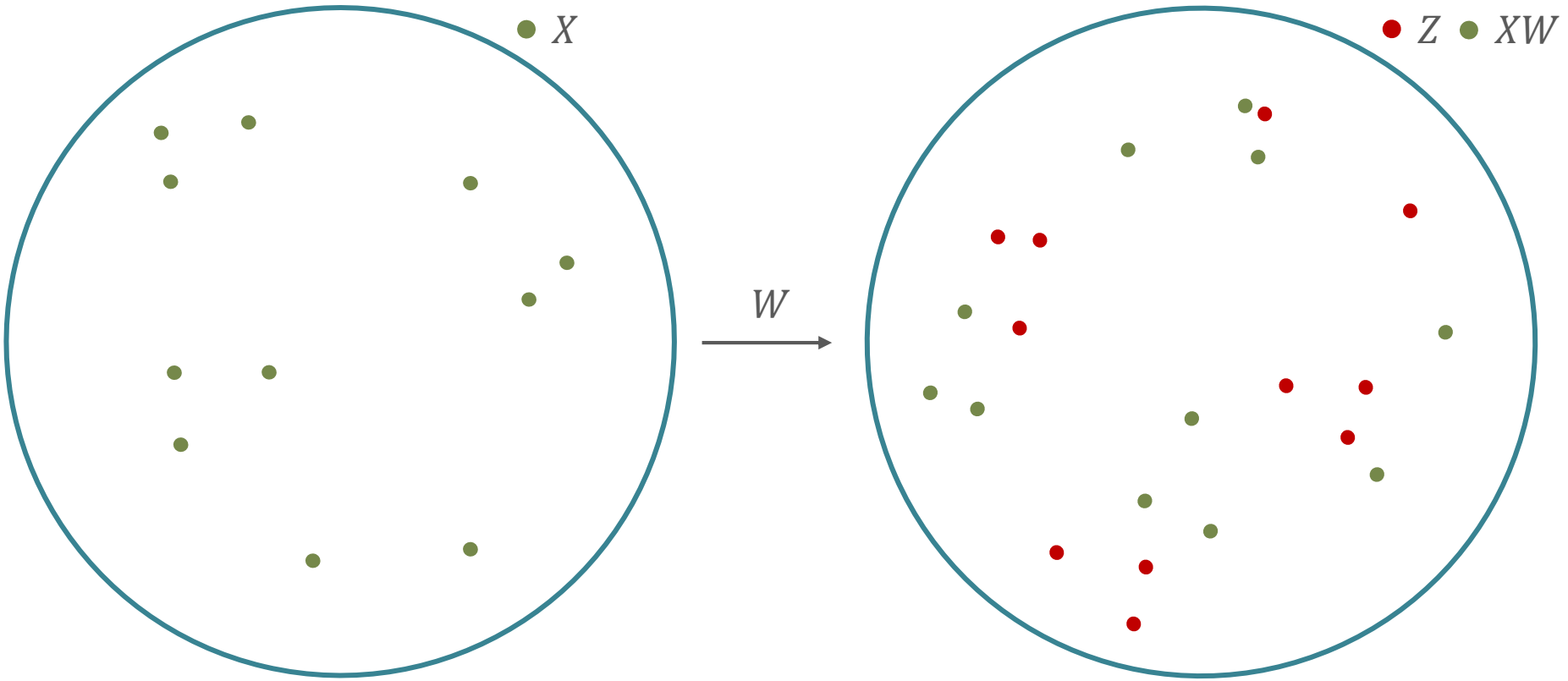
Cross-lingual embedding mappings



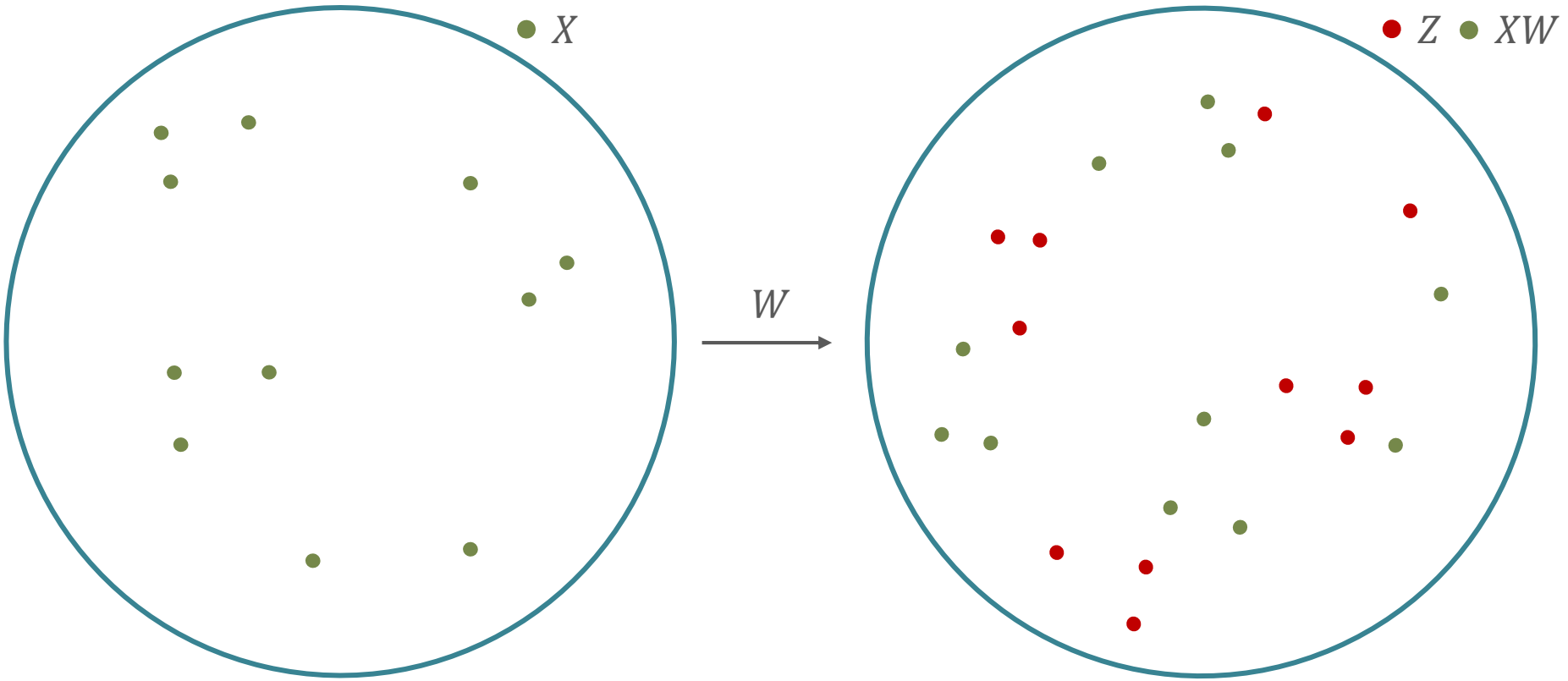
Cross-lingual embedding mappings



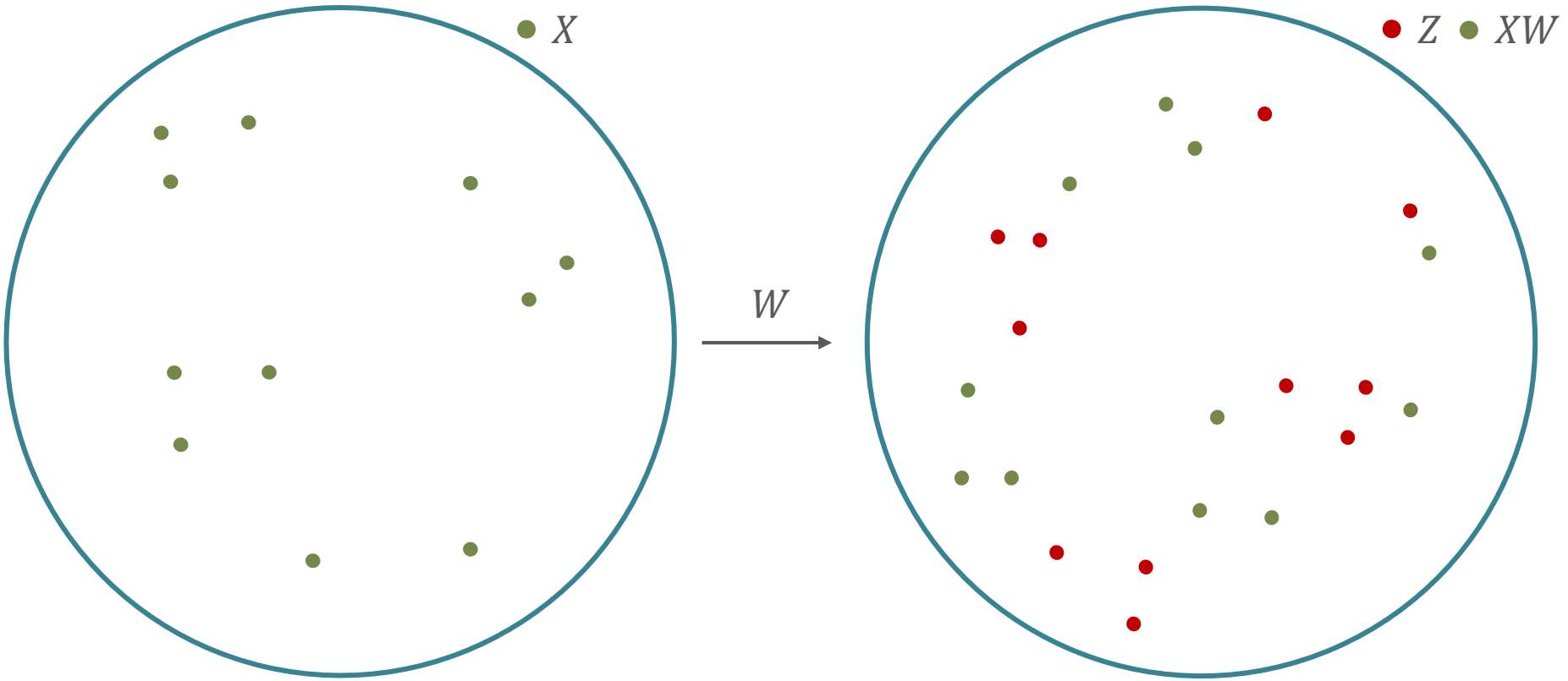
Cross-lingual embedding mappings



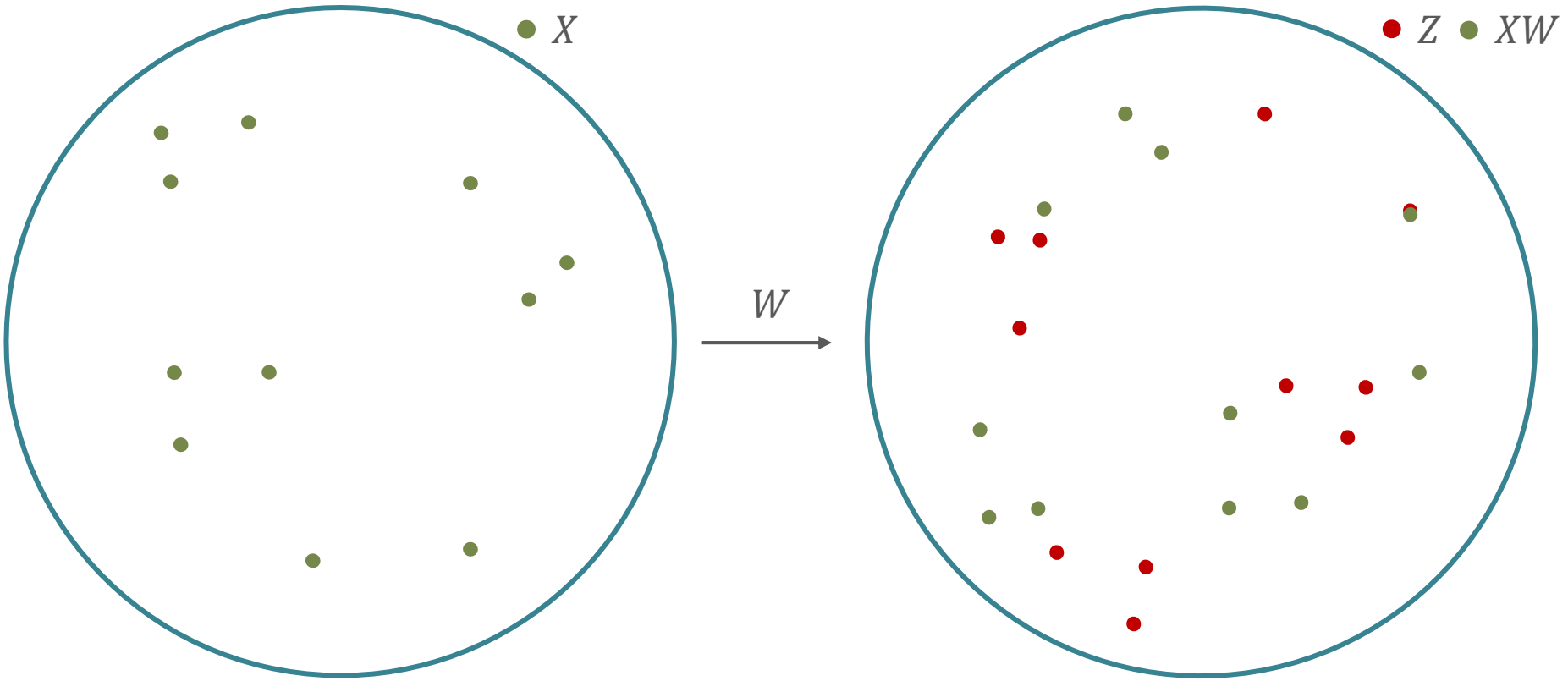
Cross-lingual embedding mappings



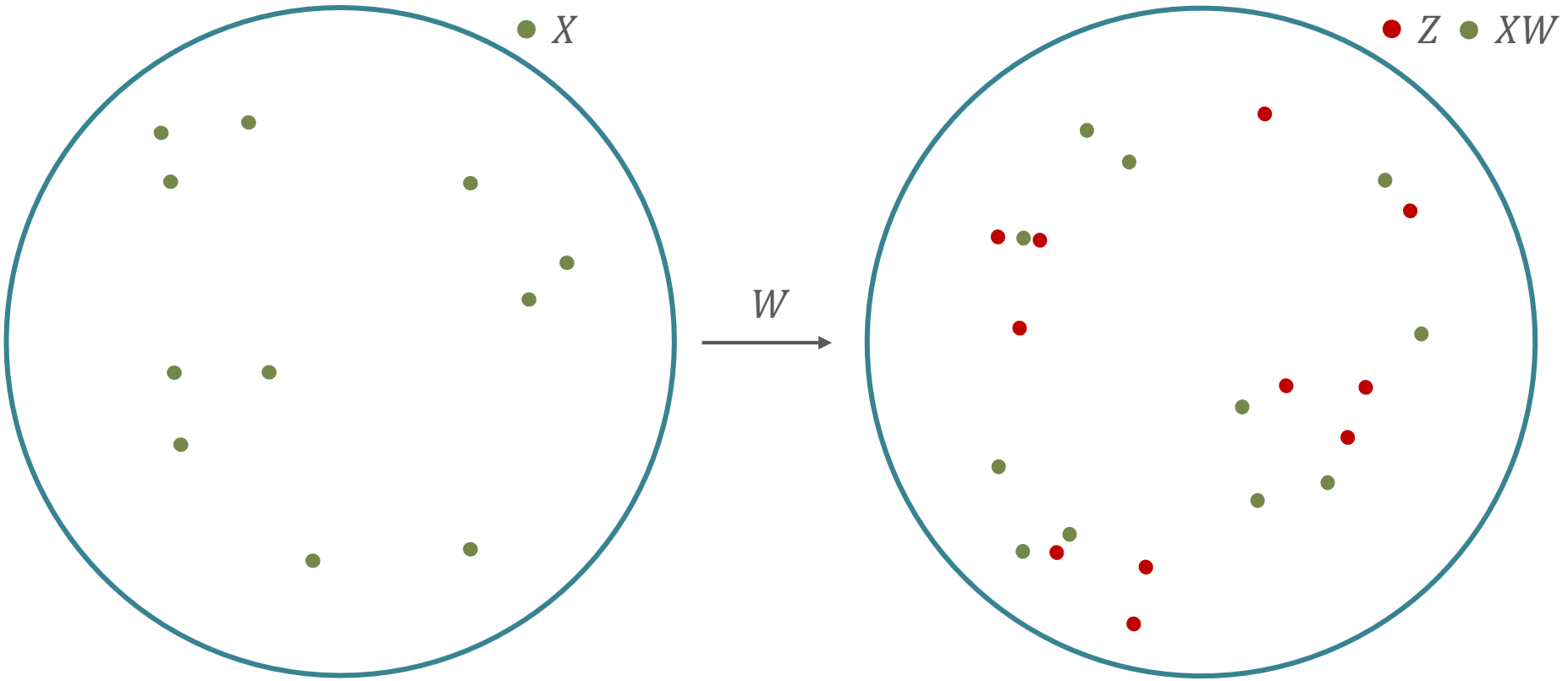
Cross-lingual embedding mappings



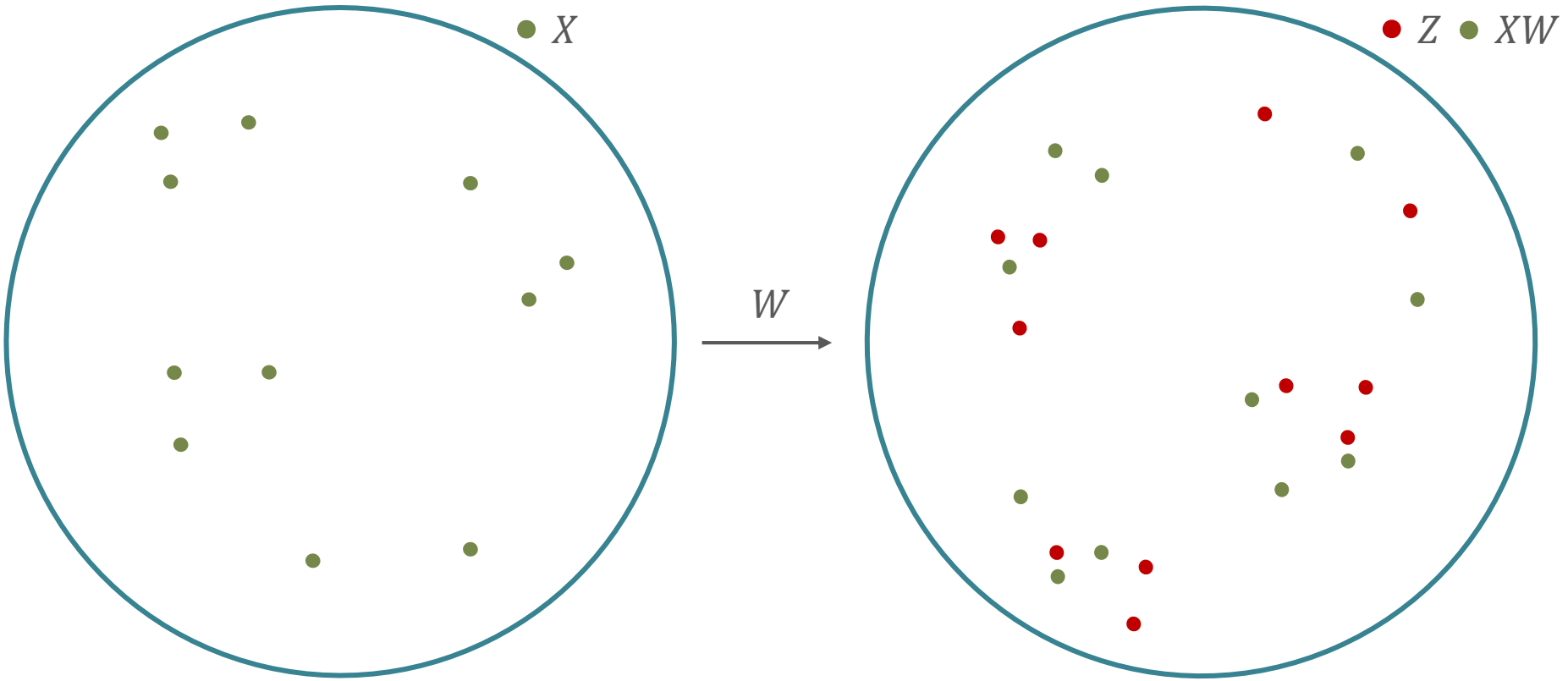
Cross-lingual embedding mappings



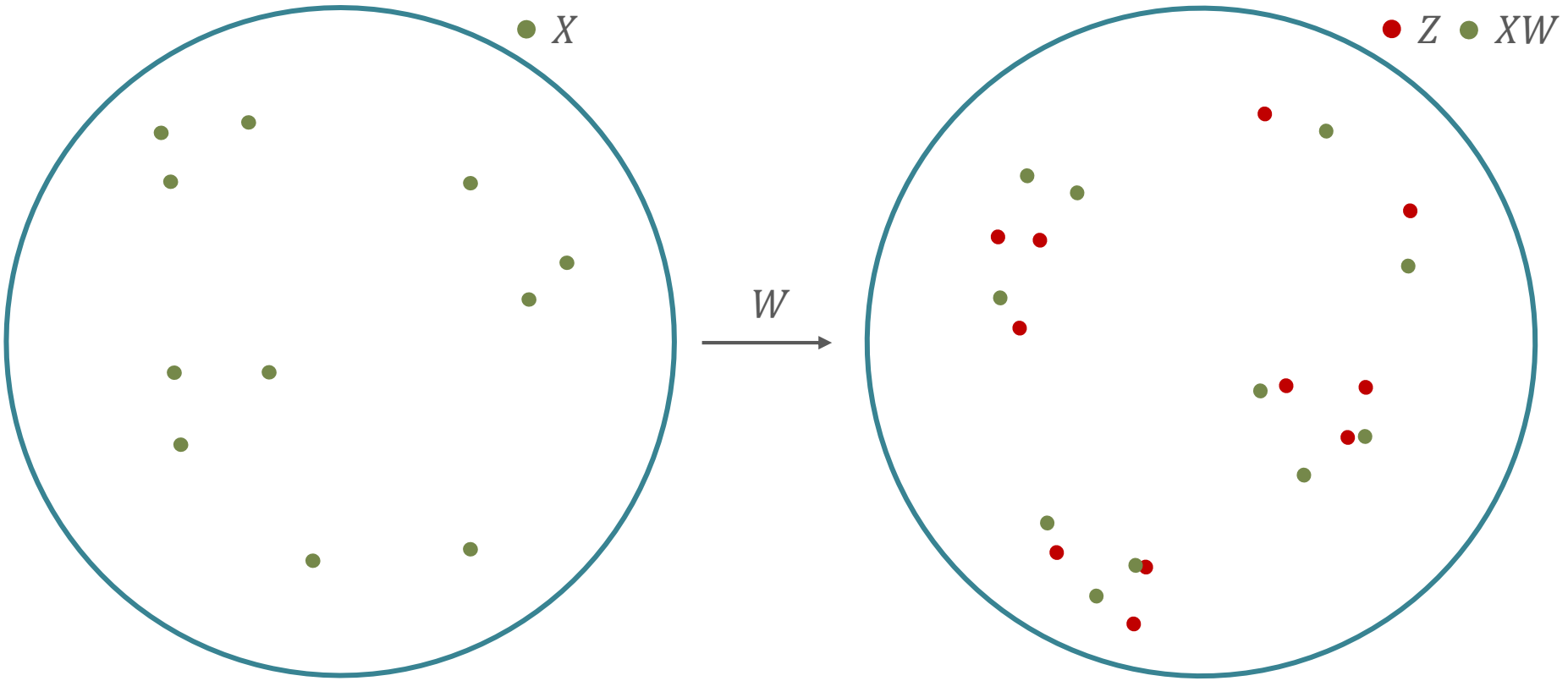
Cross-lingual embedding mappings



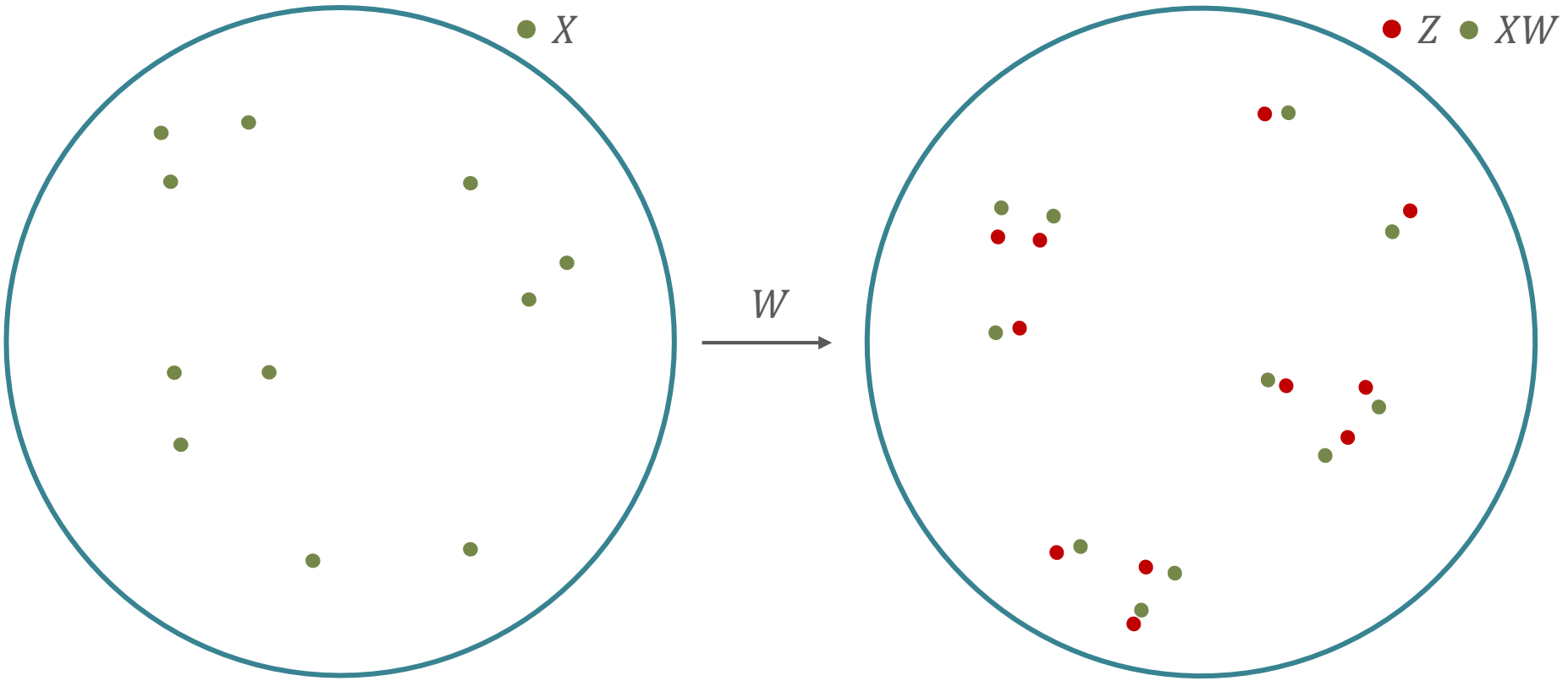
Cross-lingual embedding mappings



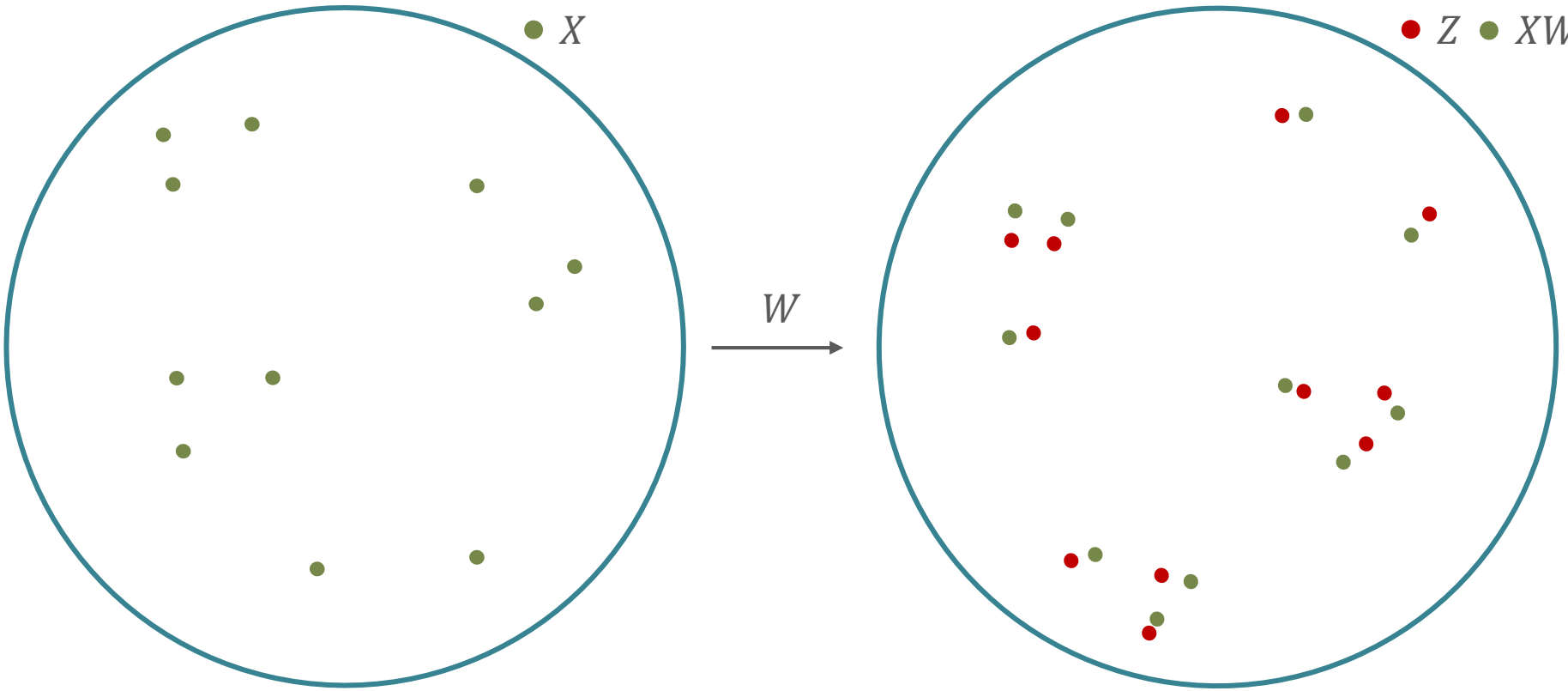
Cross-lingual embedding mappings



Cross-lingual embedding mappings



Cross-lingual embedding mappings



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

Dictionary

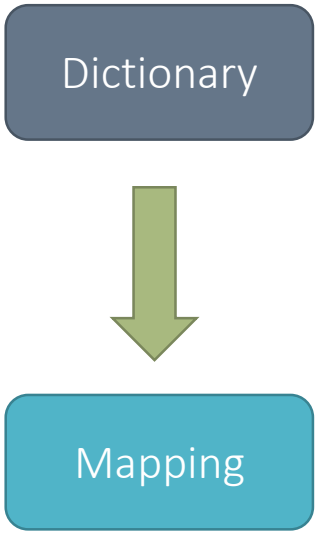
$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

Dictionary

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

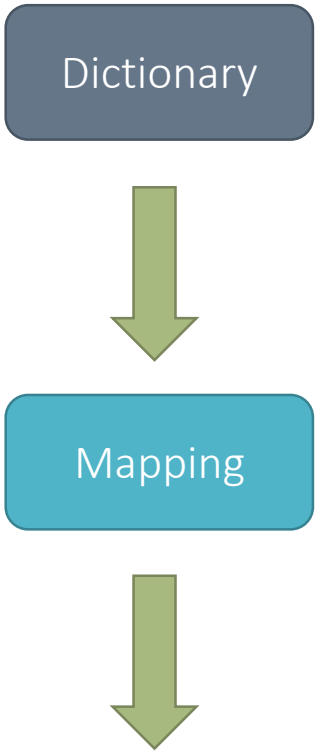
Artetxe et al. (ACL 2017)



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

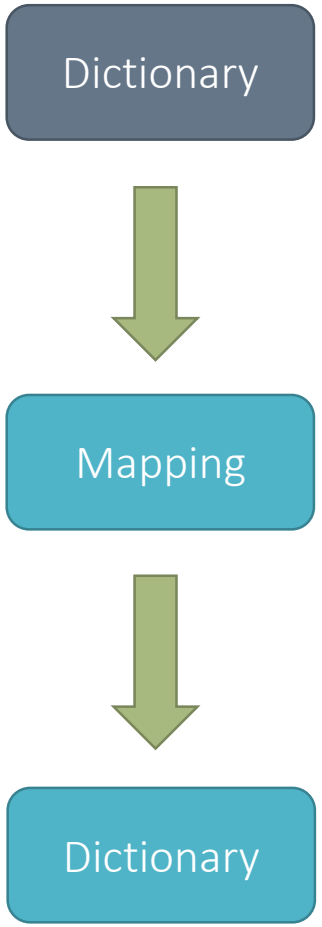
Artetxe et al. (ACL 2017)

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



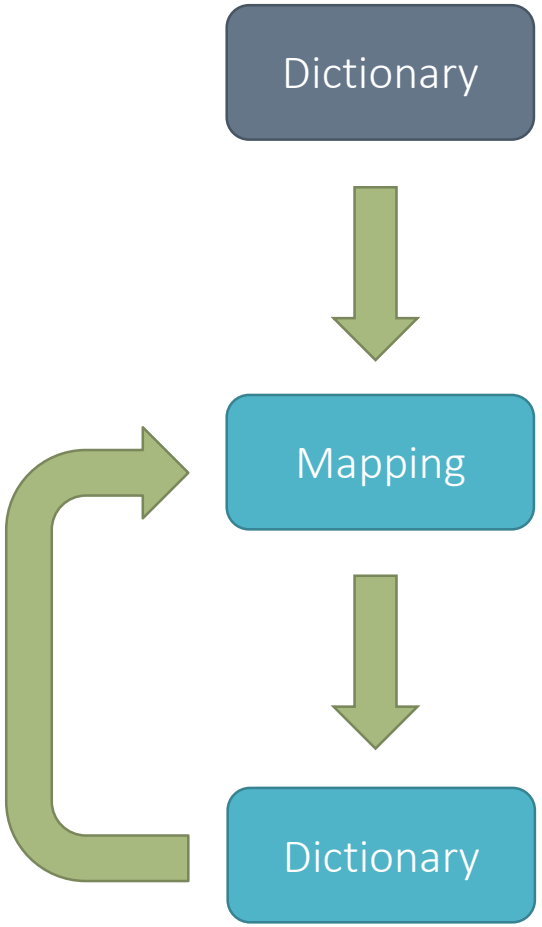
Artetxe et al. (ACL 2017)

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Artetxe et al. (ACL 2017)

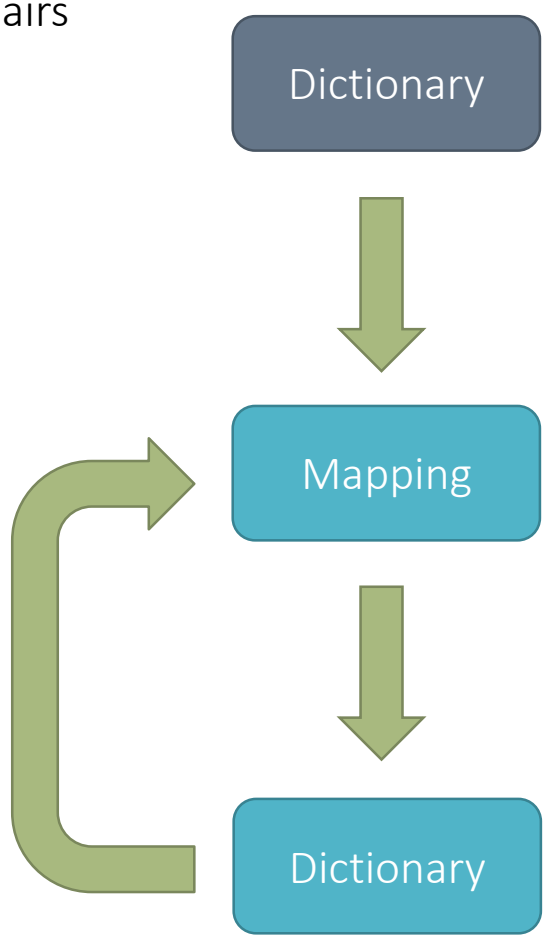
$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Artetxe et al. (ACL 2017)

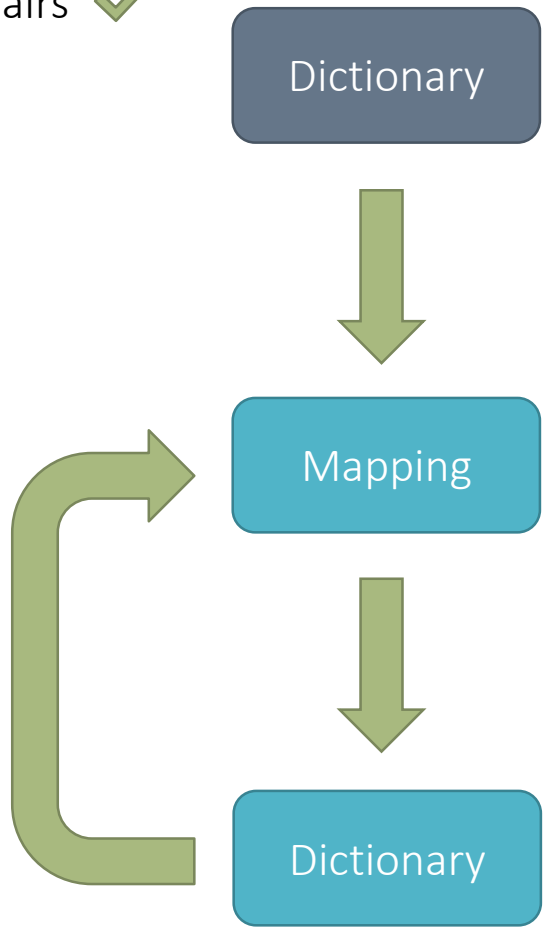
- 25 word pairs

$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Artetxe et al. (ACL 2017)

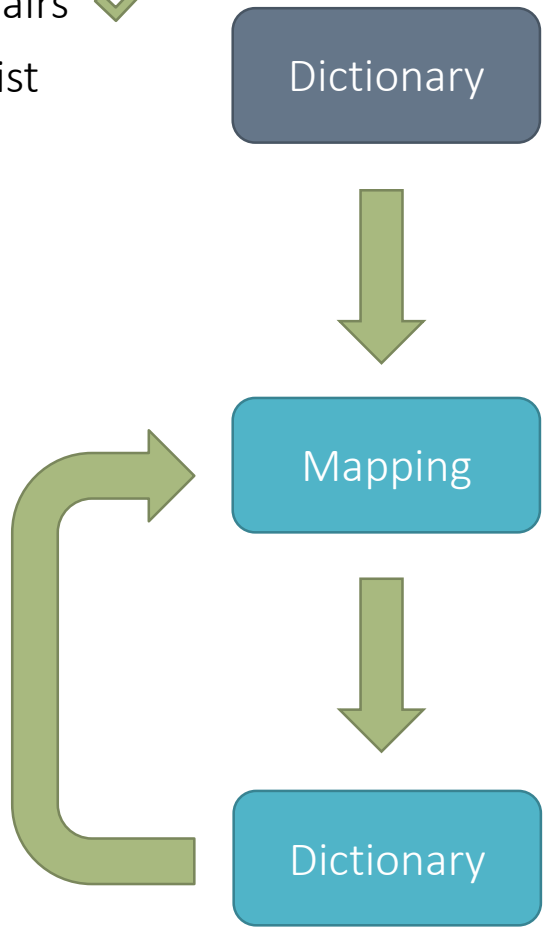
- 25 word pairs ✓



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

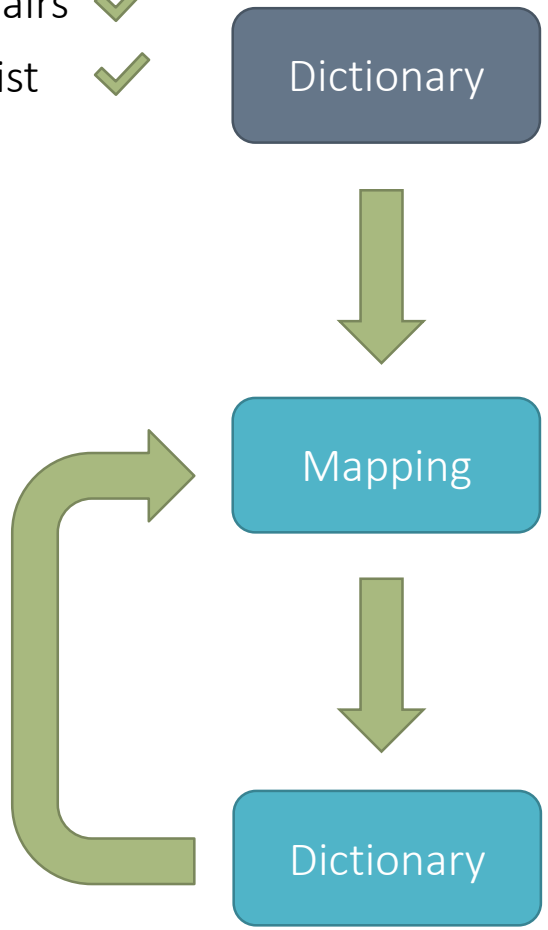
- 25 word pairs ✓
- Numeral list



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

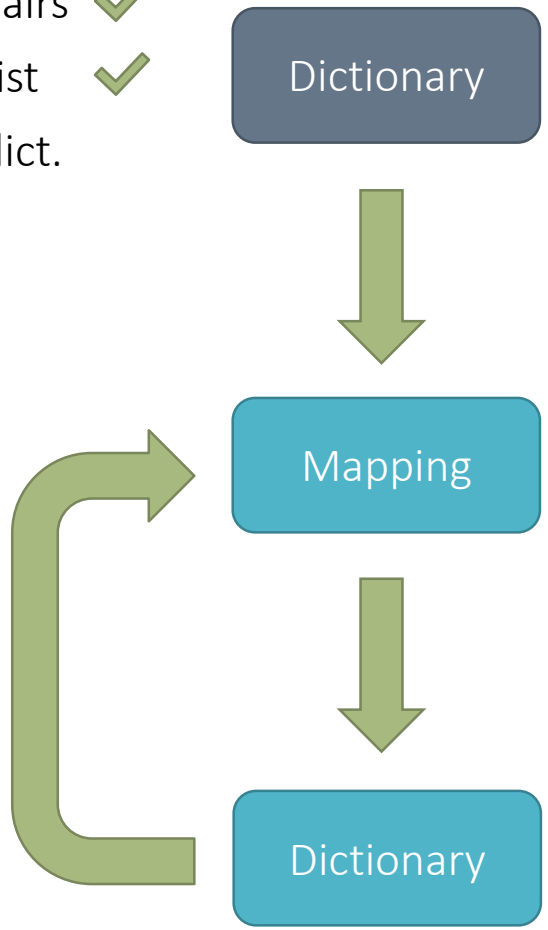
- 25 word pairs ✓
- Numeral list ✓



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

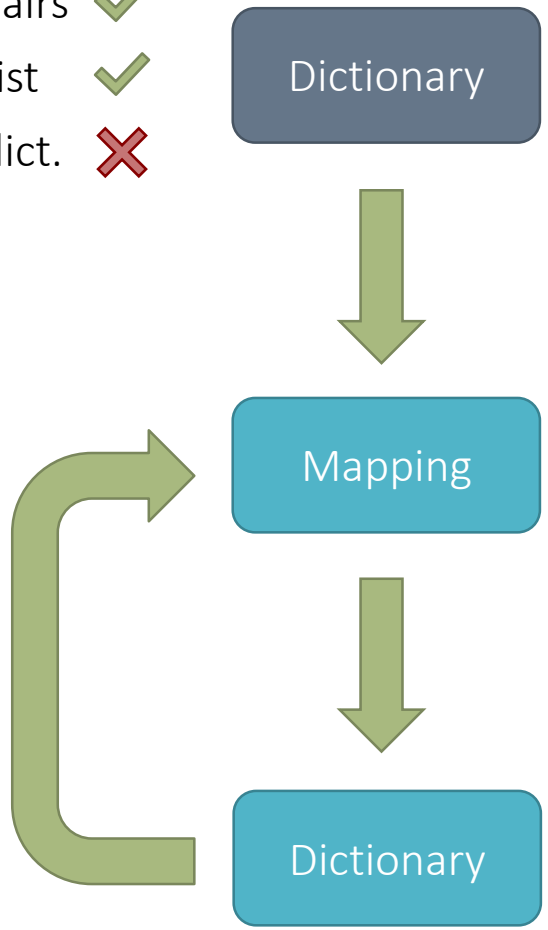
- 25 word pairs ✓
- Numeral list ✓
- Random dict.



$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

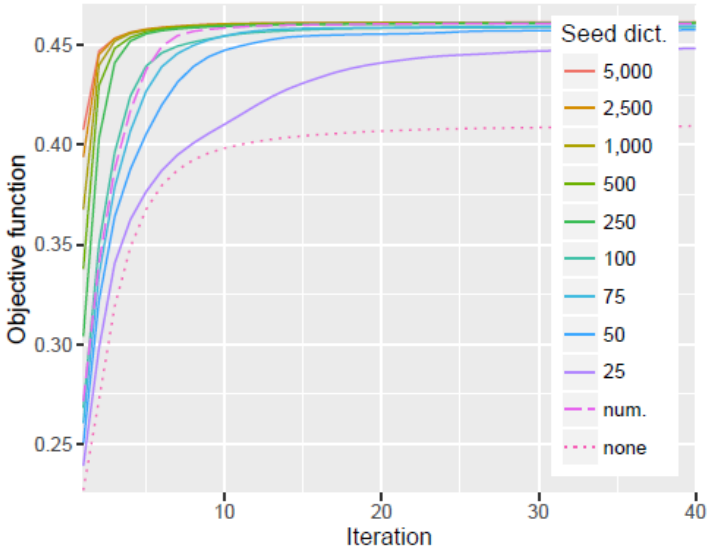
Artetxe et al. (ACL 2017)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗

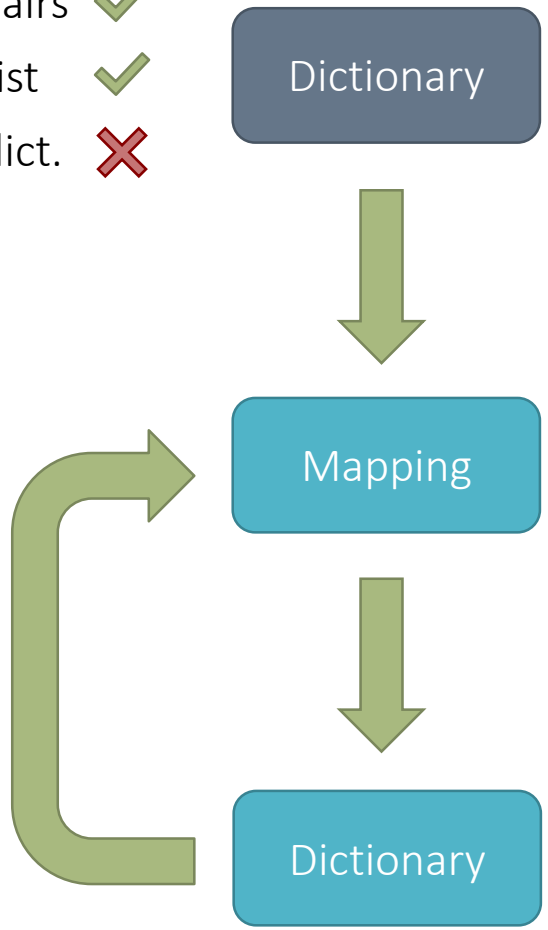


$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

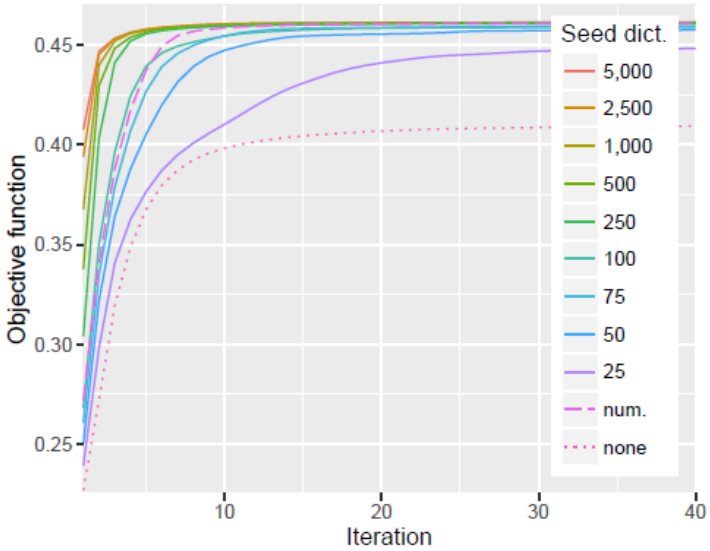


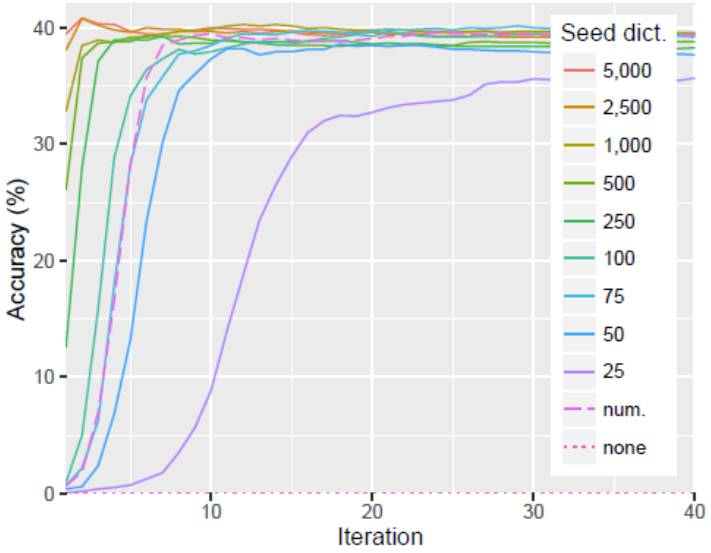
- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



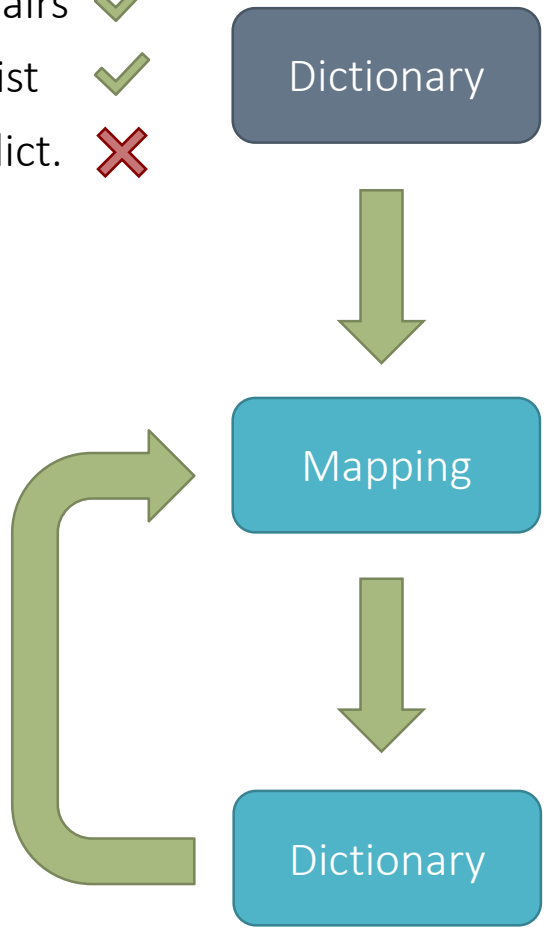
$$W^* = \arg\min_{W \in O(n)} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Artetxe et al. (ACL 2017)

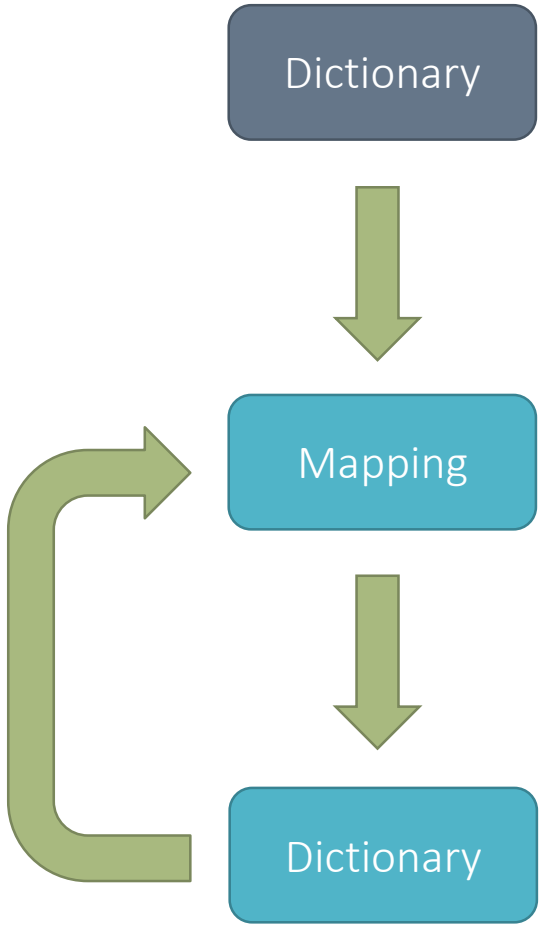




- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗

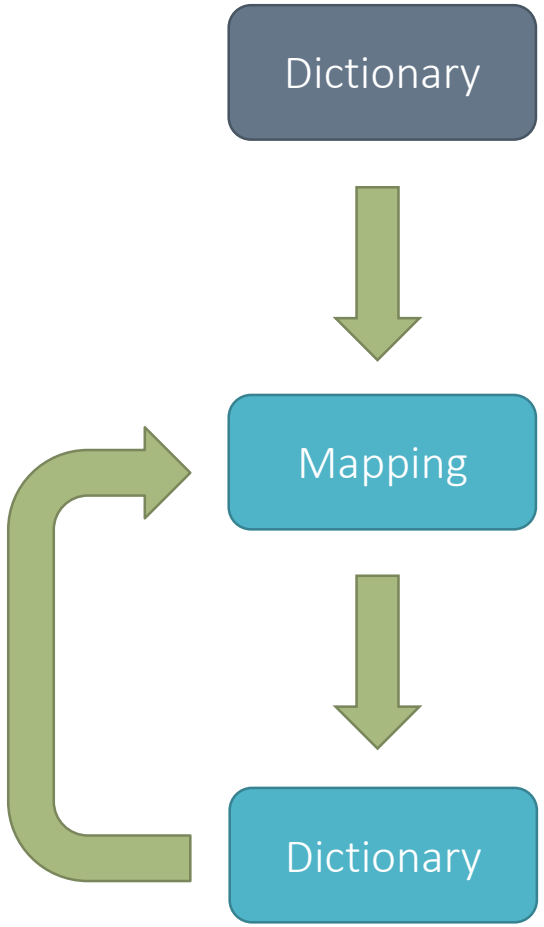


Proposed method



Proposed method

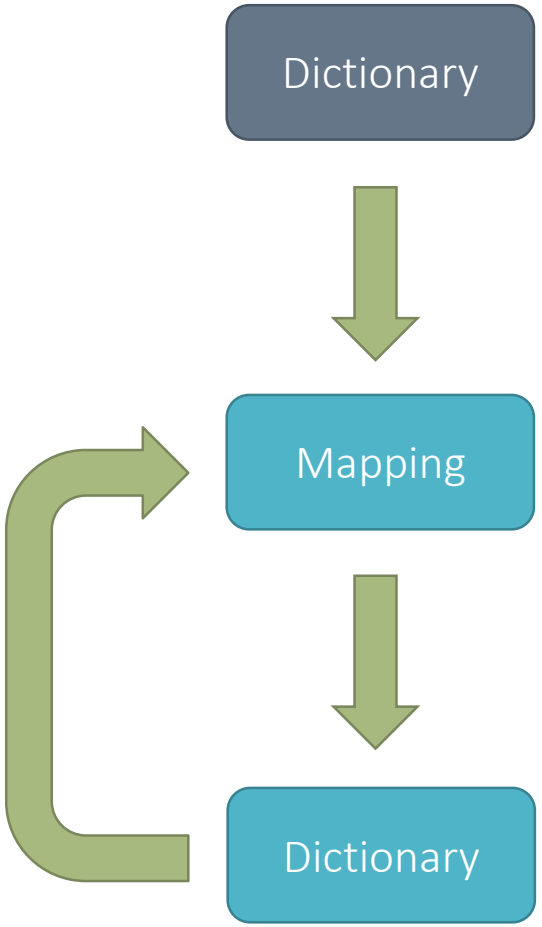
1) Fully unsupervised initialization



Proposed method

1) Fully unsupervised initialization

2) Robust self-learning

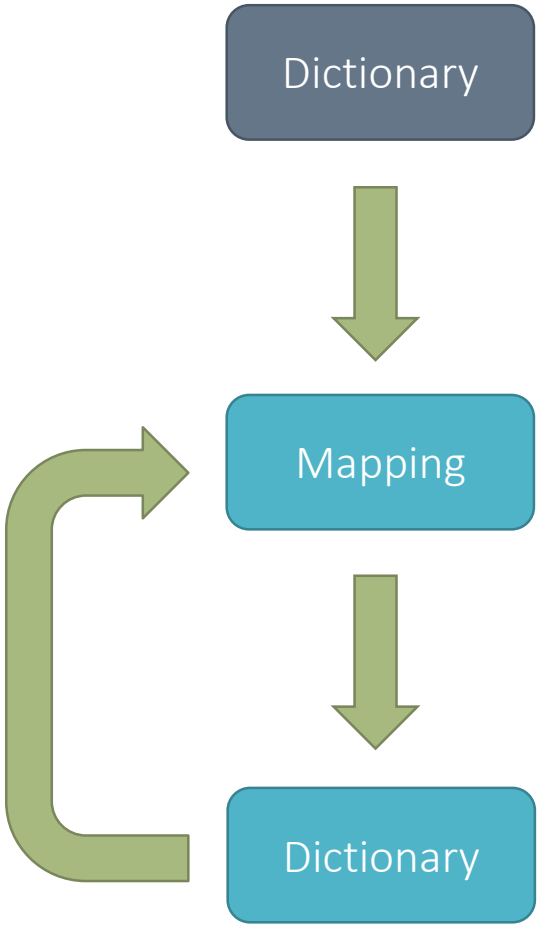


Proposed method

1) Fully unsupervised initialization

Intra-lingual similarity distribution

2) Robust self-learning

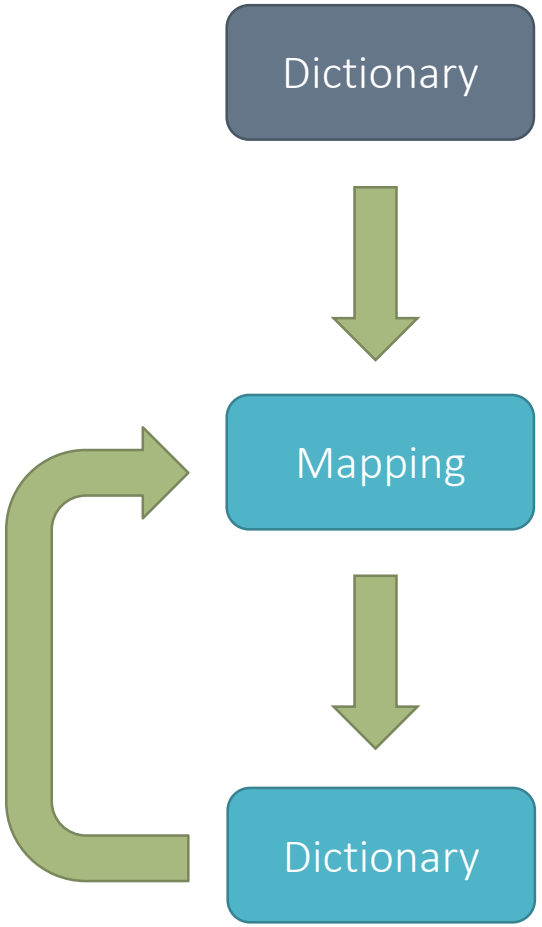


Proposed method

1) Fully unsupervised initialization

English

Intra-lingual similarity distribution



2) Robust self-learning

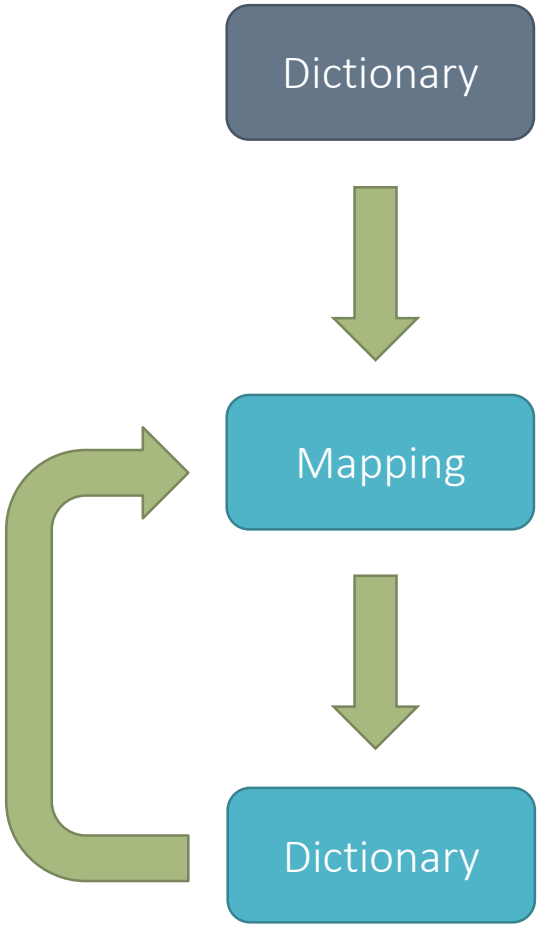
Proposed method

1) Fully unsupervised initialization

English

two

Intra-lingual similarity distribution



2) Robust self-learning

Proposed method

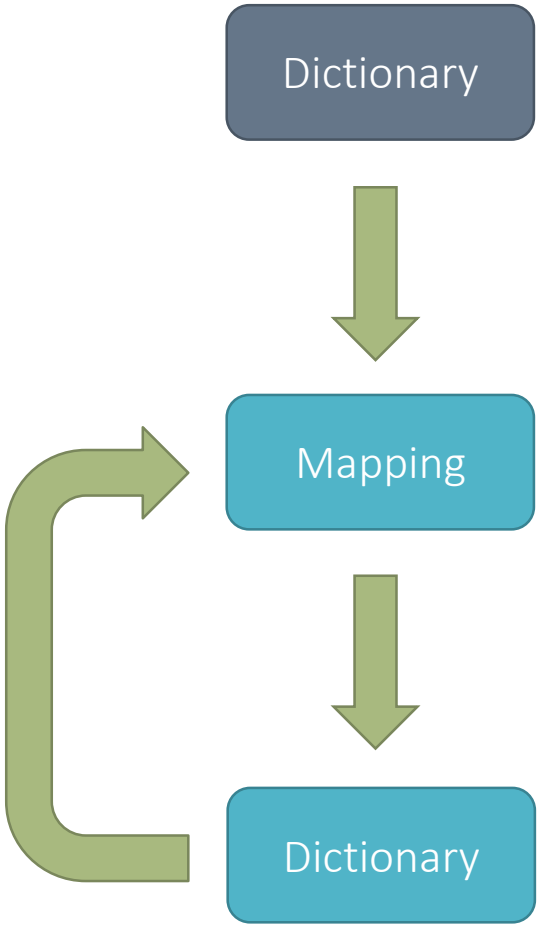
1) Fully unsupervised initialization

English

```
for x in vocab:
   sim("two", x)
```

two

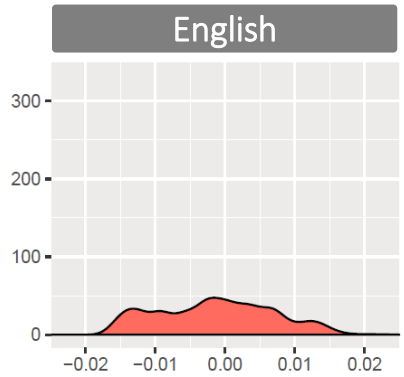
Intra-lingual similarity distribution



2) Robust self-learning

Proposed method

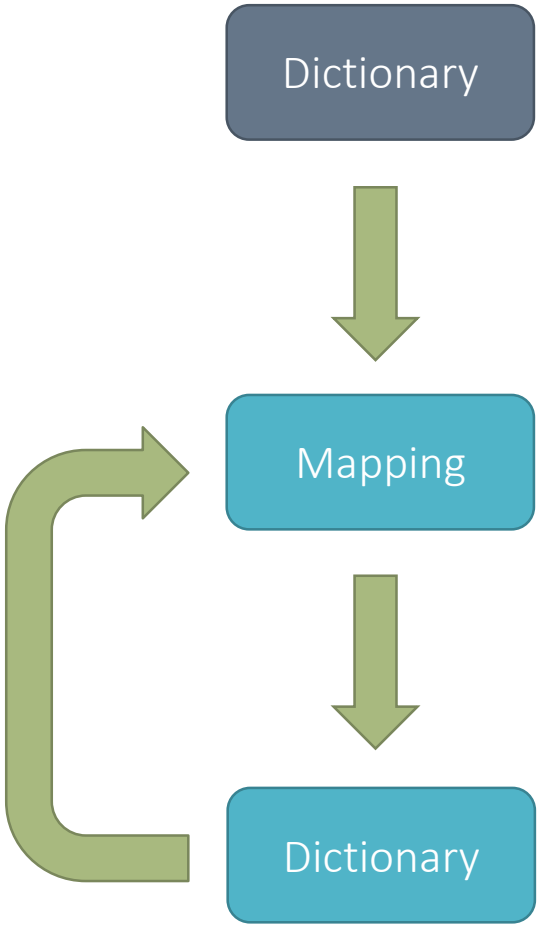
1) Fully unsupervised initialization



two

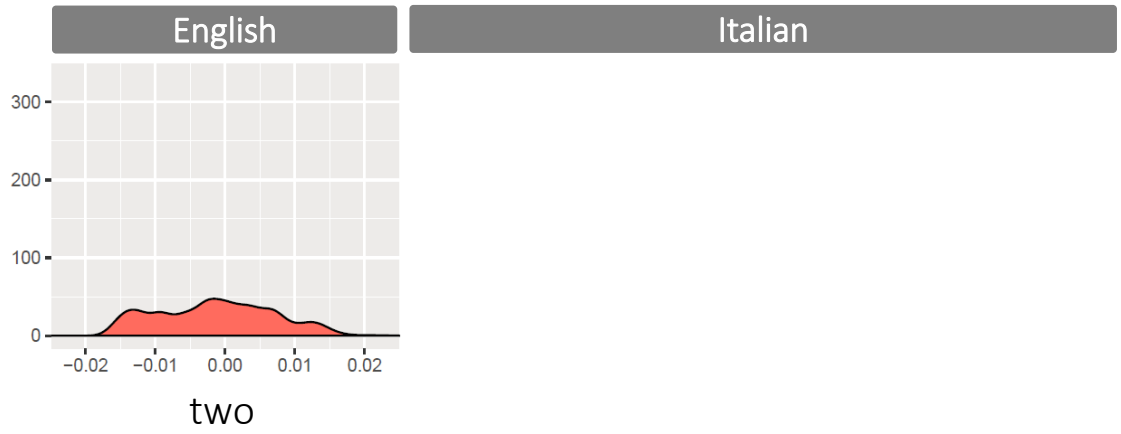
Intra-lingual similarity distribution

2) Robust self-learning



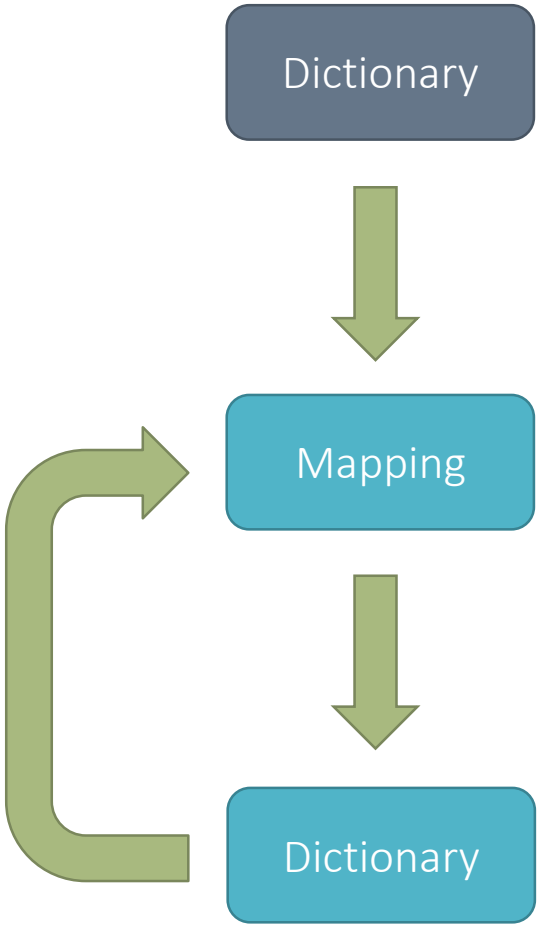
Proposed method

1) Fully unsupervised initialization



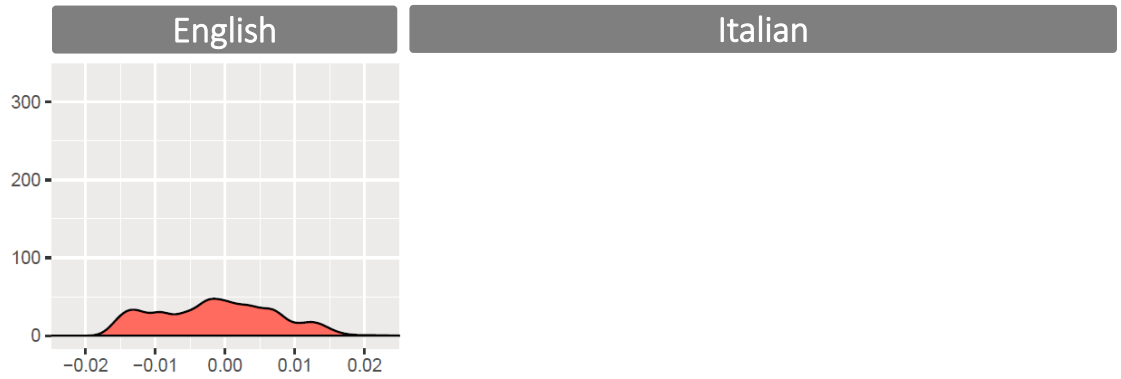
Intra-lingual similarity distribution

2) Robust self-learning



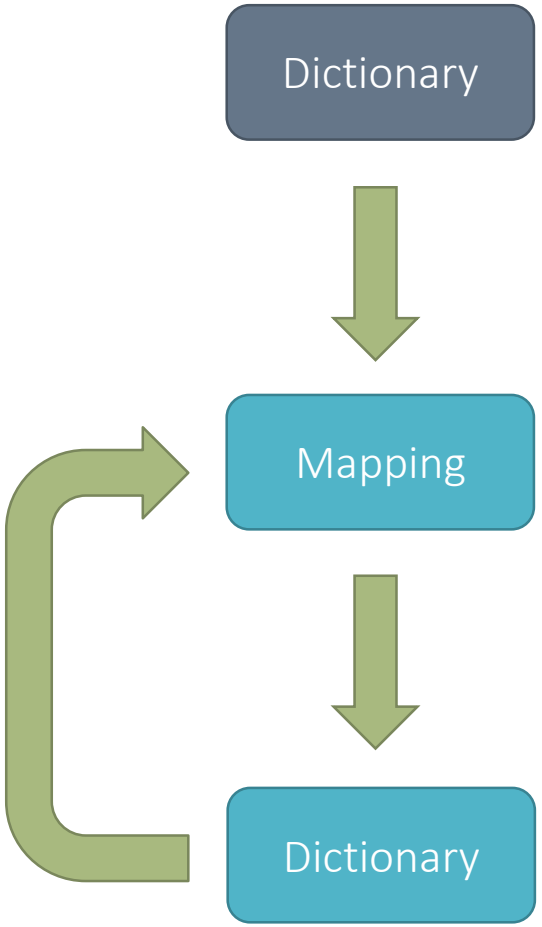
Proposed method

1) Fully unsupervised initialization



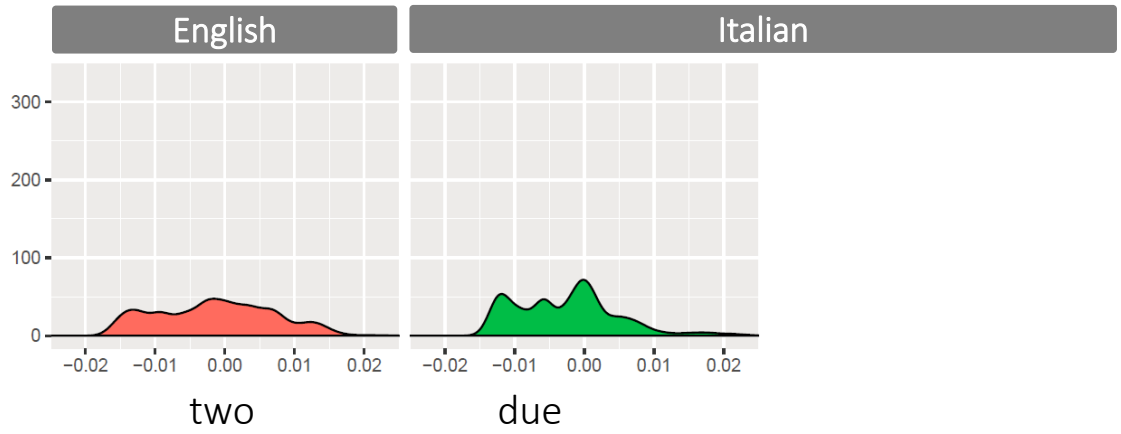
Intra-lingual similarity distribution

2) Robust self-learning



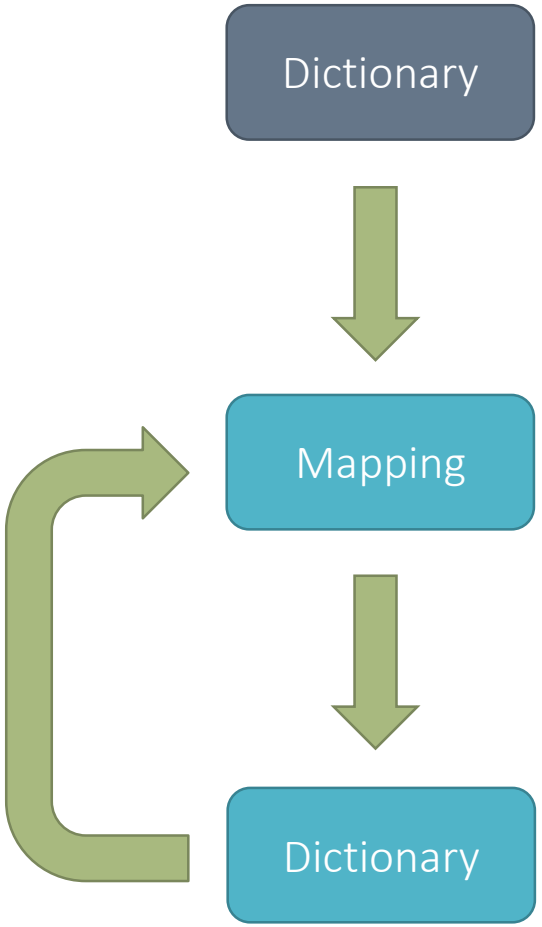
Proposed method

1) Fully unsupervised initialization



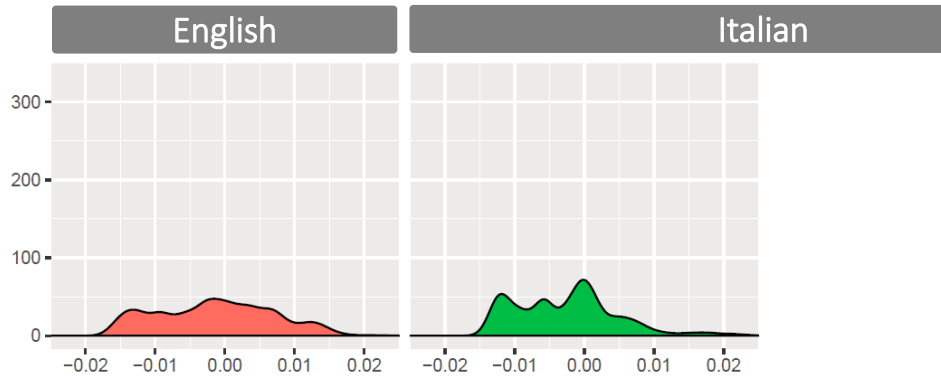
Intra-lingual similarity distribution

2) Robust self-learning



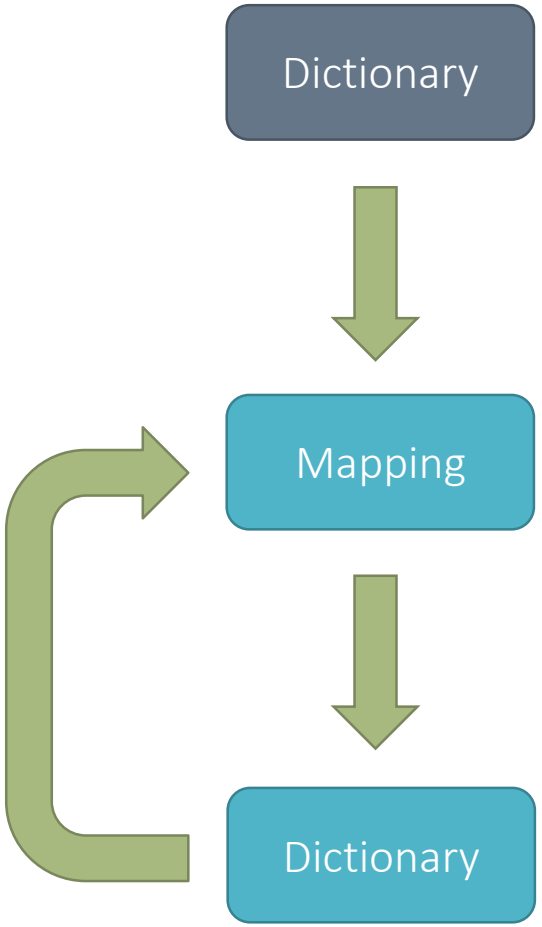
Proposed method

1) Fully unsupervised initialization



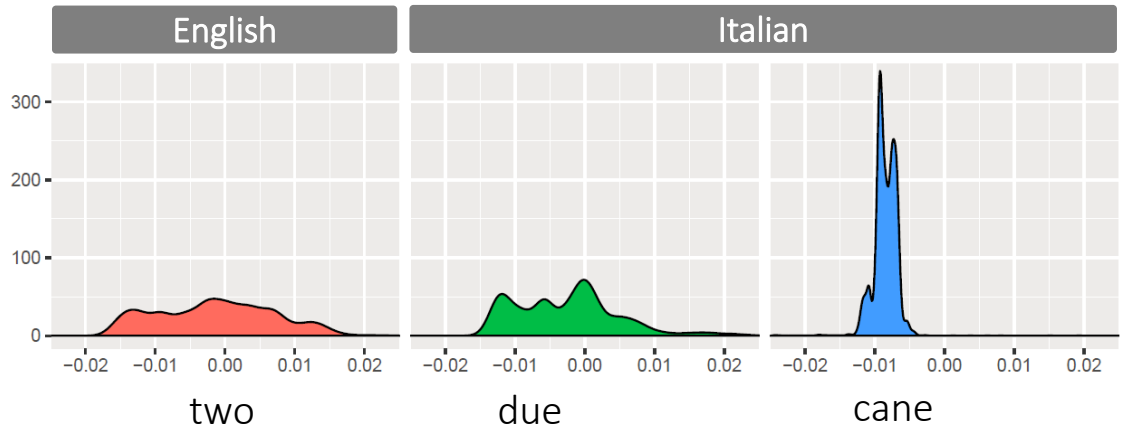
Intra-lingual similarity distribution

2) Robust self-learning



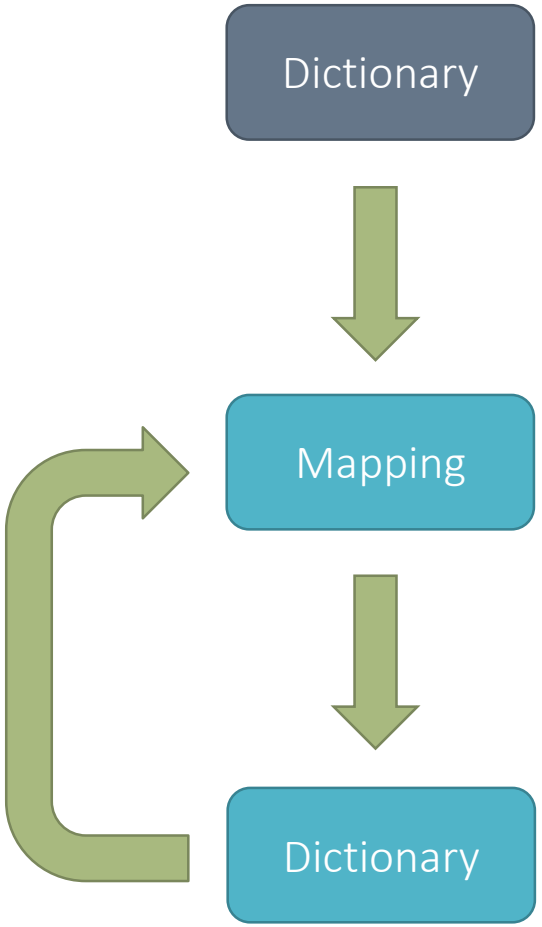
Proposed method

1) Fully unsupervised initialization



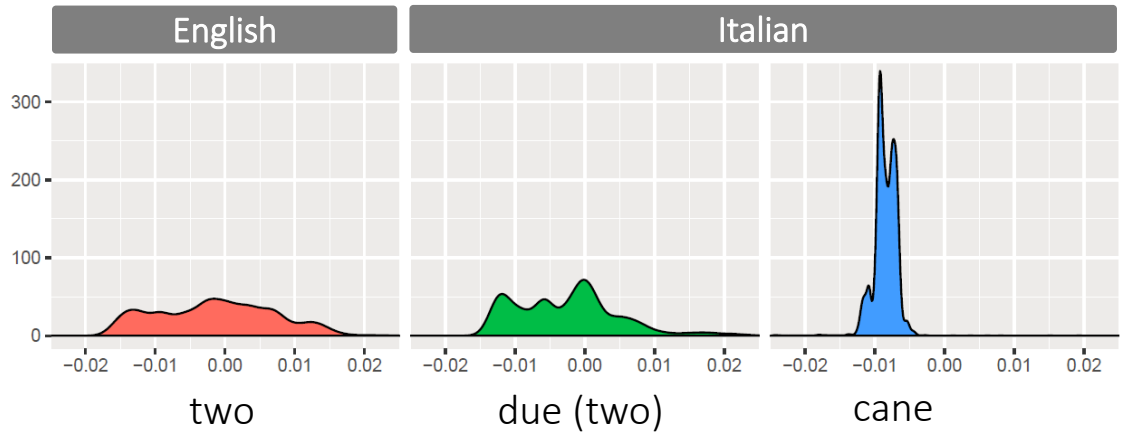
Intra-lingual similarity distribution

2) Robust self-learning



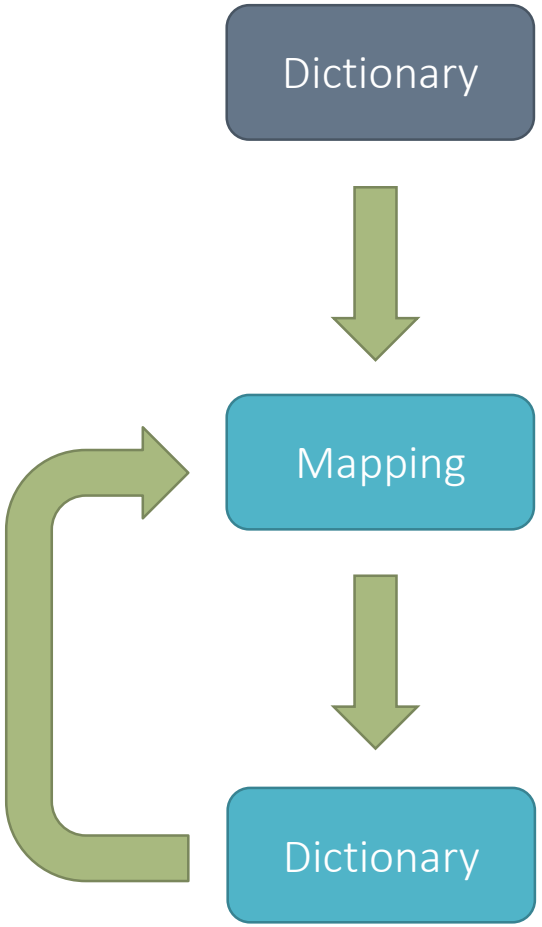
Proposed method

1) Fully unsupervised initialization



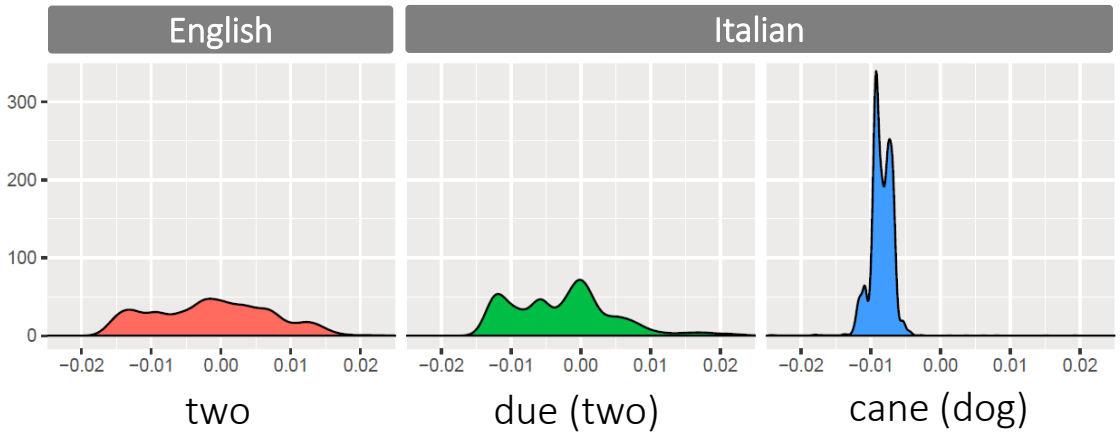
Intra-lingual similarity distribution

2) Robust self-learning



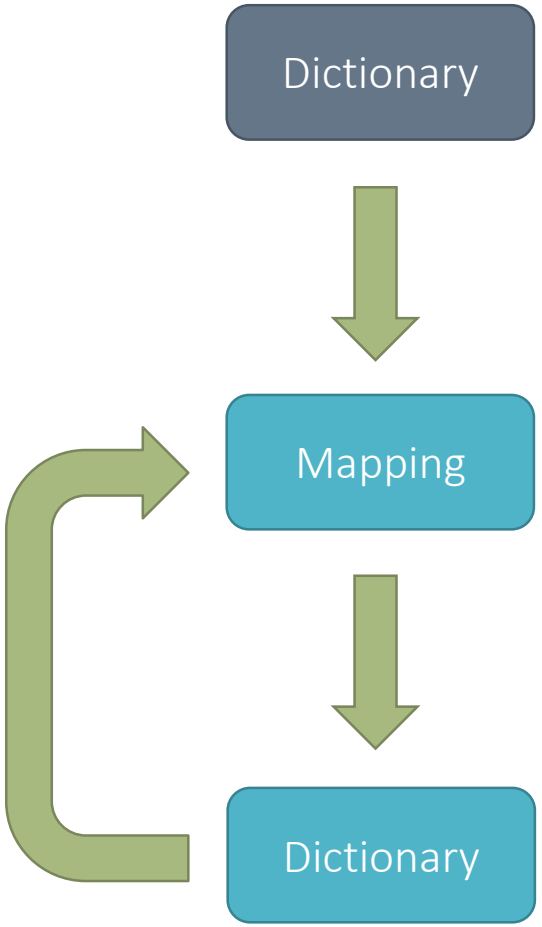
Proposed method

1) Fully unsupervised initialization



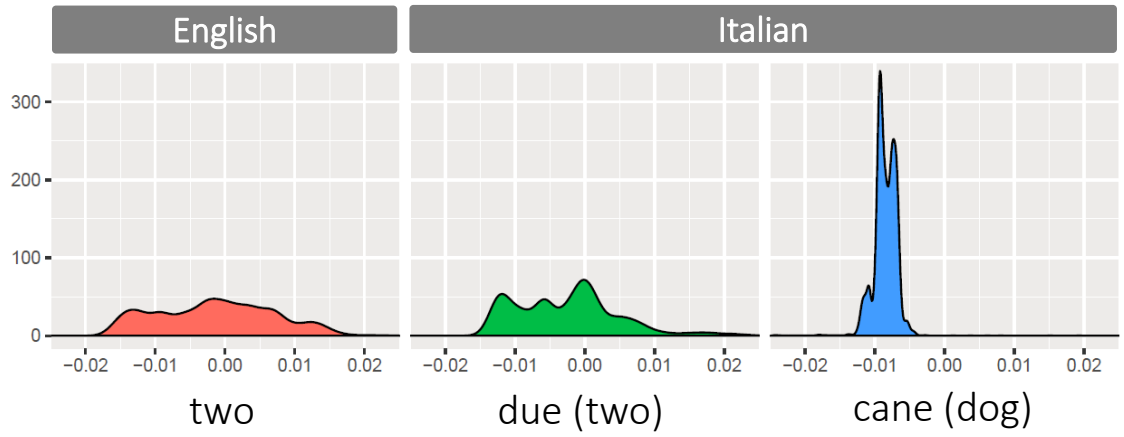
Intra-lingual similarity distribution

2) Robust self-learning

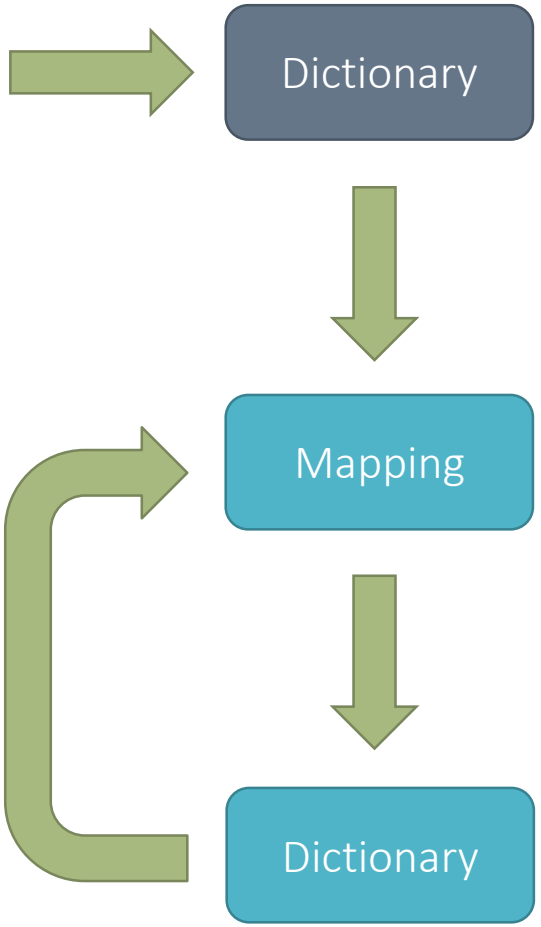


Proposed method

1) Fully unsupervised initialization



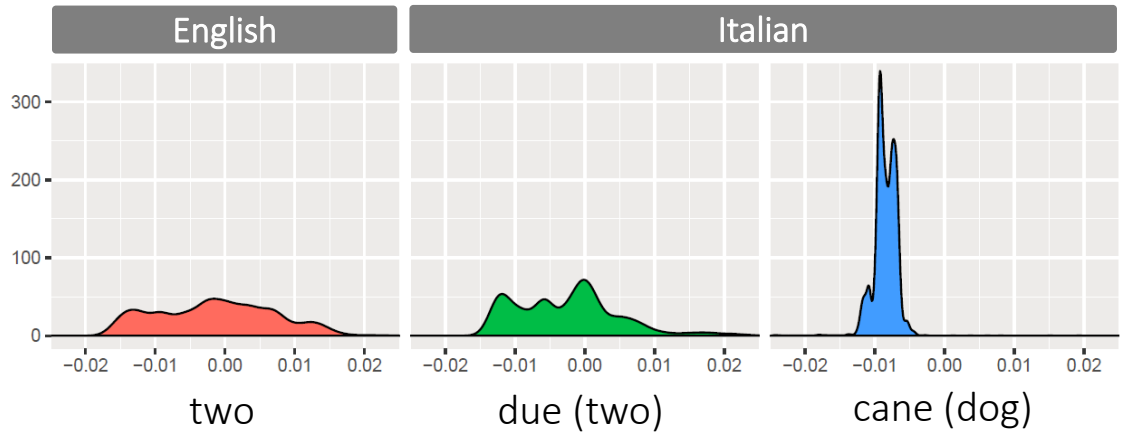
Intra-lingual similarity distribution



2) Robust self-learning

Proposed method

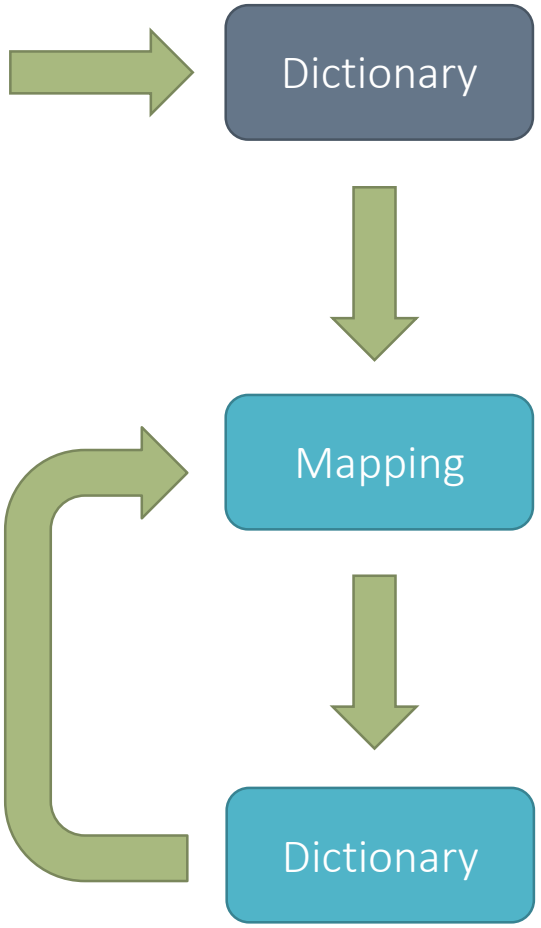
1) Fully unsupervised initialization



Intra-lingual similarity distribution

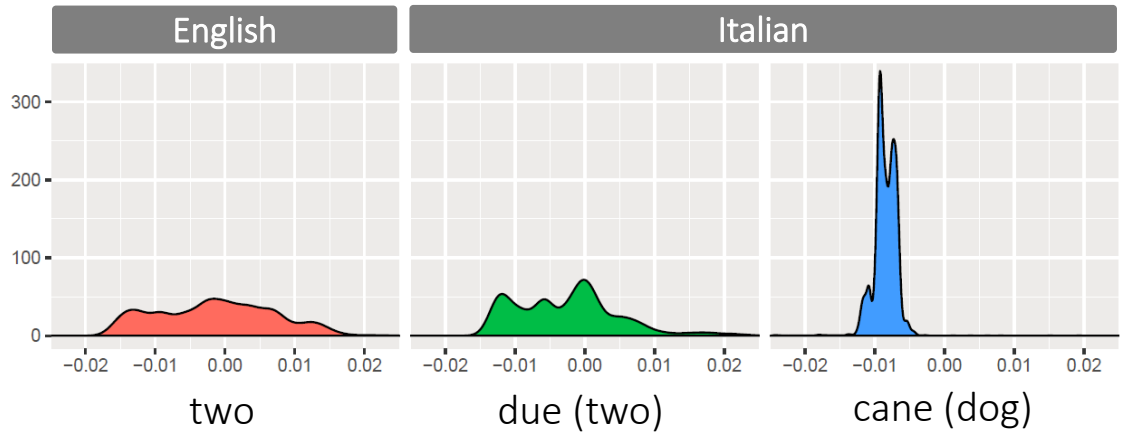
$$X' = \text{sorted}\left(\sqrt{XX^T}\right)$$

2) Robust self-learning



Proposed method

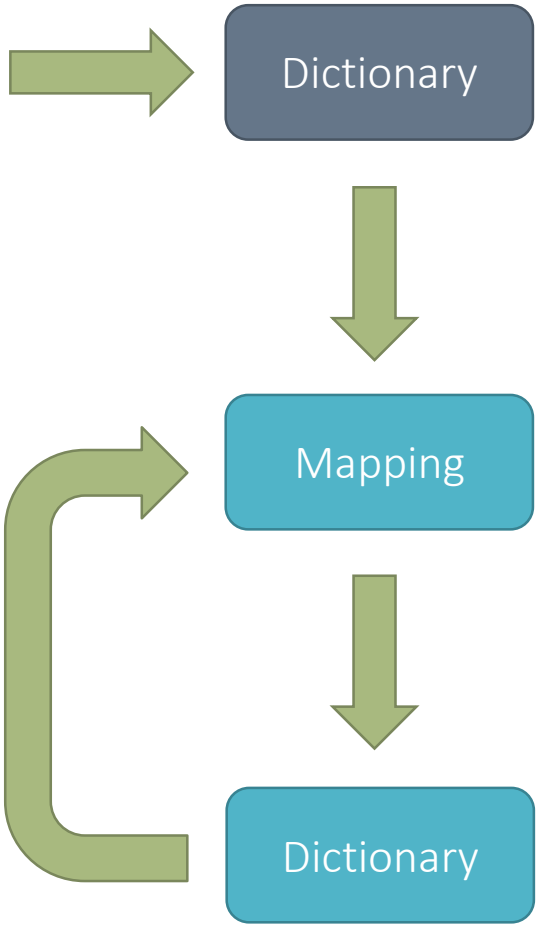
1) Fully unsupervised initialization



Intra-lingual similarity distribution

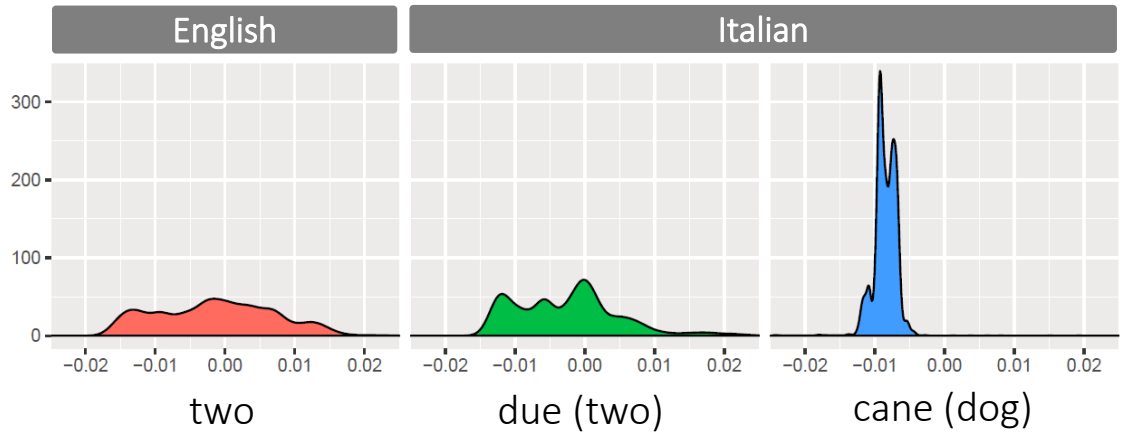
$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

2) Robust self-learning



Proposed method

1) Fully unsupervised initialization

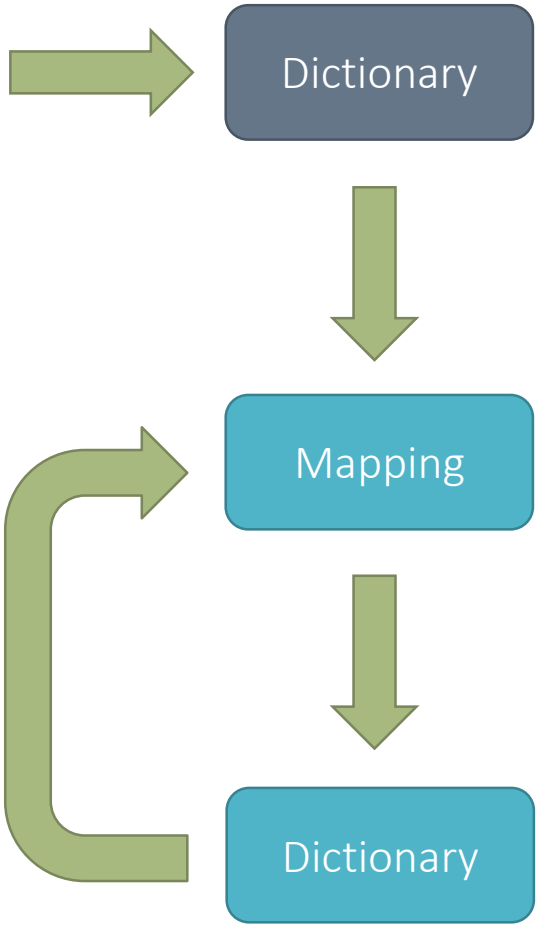


Intra-lingual similarity distribution

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

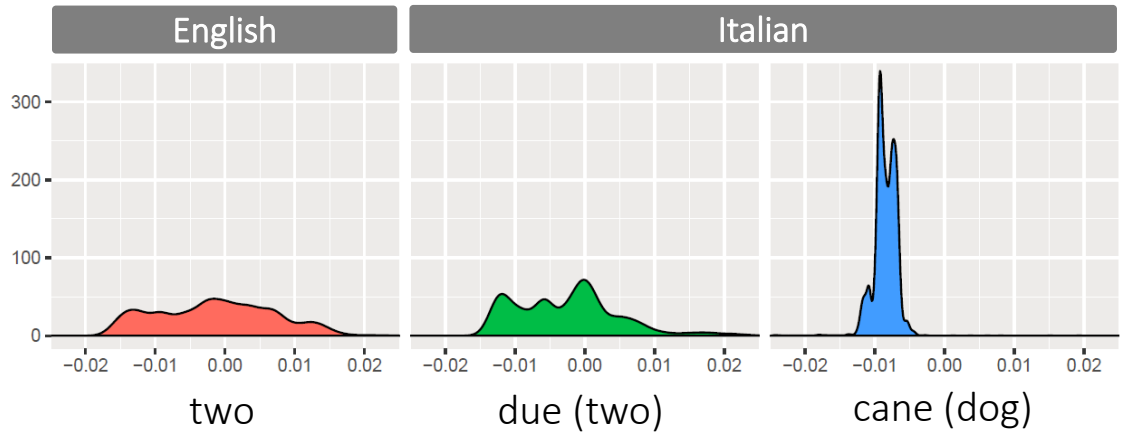
2) Robust self-learning

- Stochastic dictionary induction



Proposed method

1) Fully unsupervised initialization

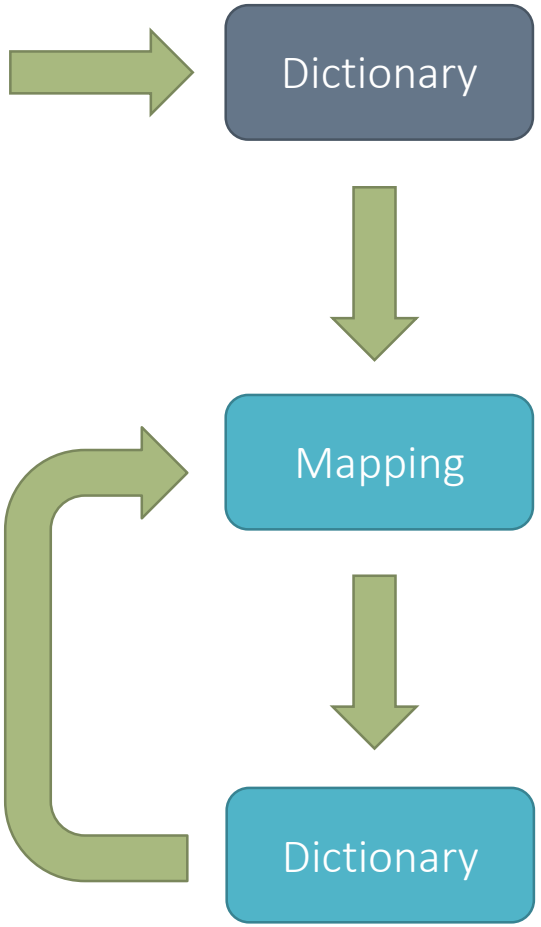


Intra-lingual similarity distribution

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

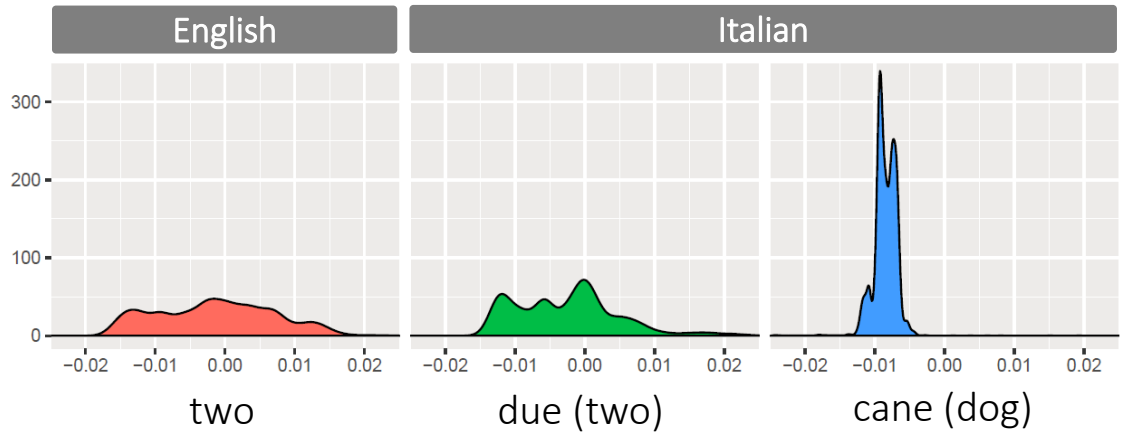
2) Robust self-learning

- Stochastic dictionary induction
- Frequency-based vocabulary cutoff



Proposed method

1) Fully unsupervised initialization

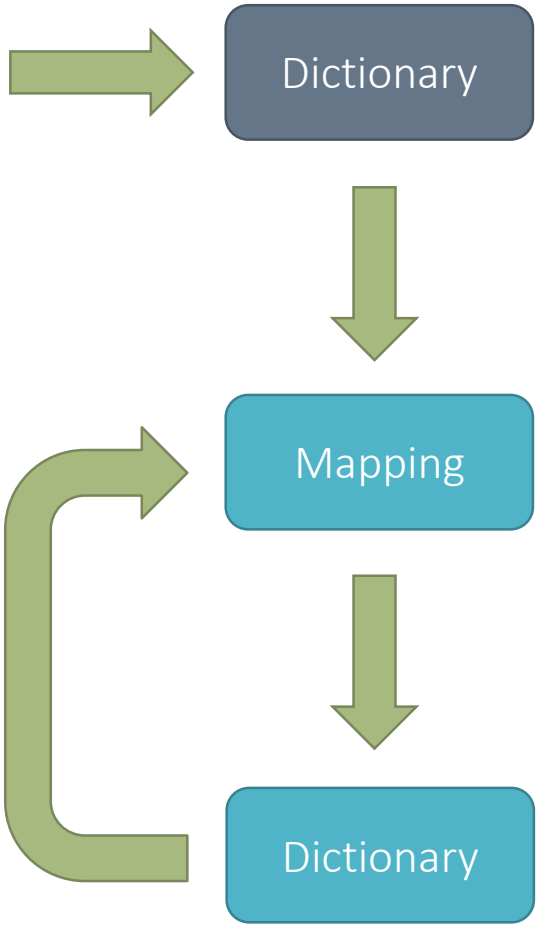


Intra-lingual similarity distribution

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

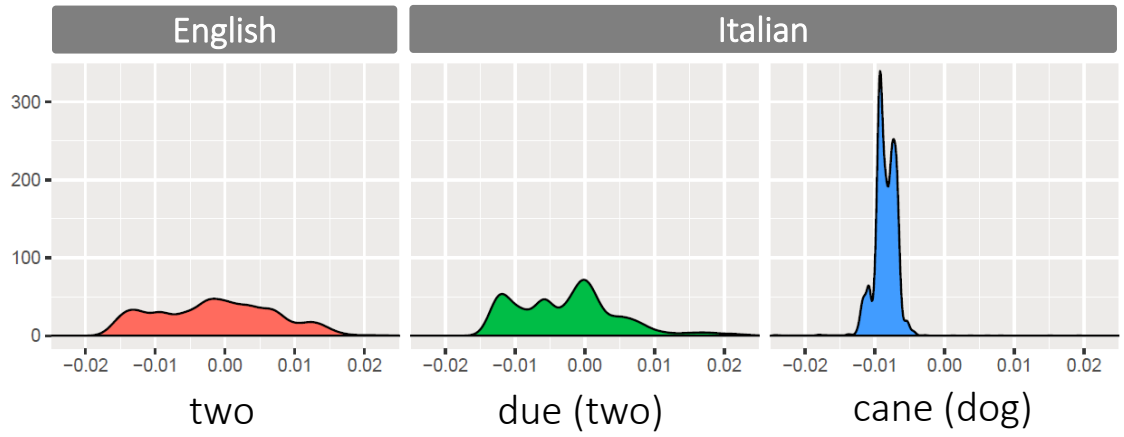
2) Robust self-learning

- Stochastic dictionary induction
- Frequency-based vocabulary cutoff
- CSLS retrieval (Conneau et al., 2018)



Proposed method

1) Fully unsupervised initialization

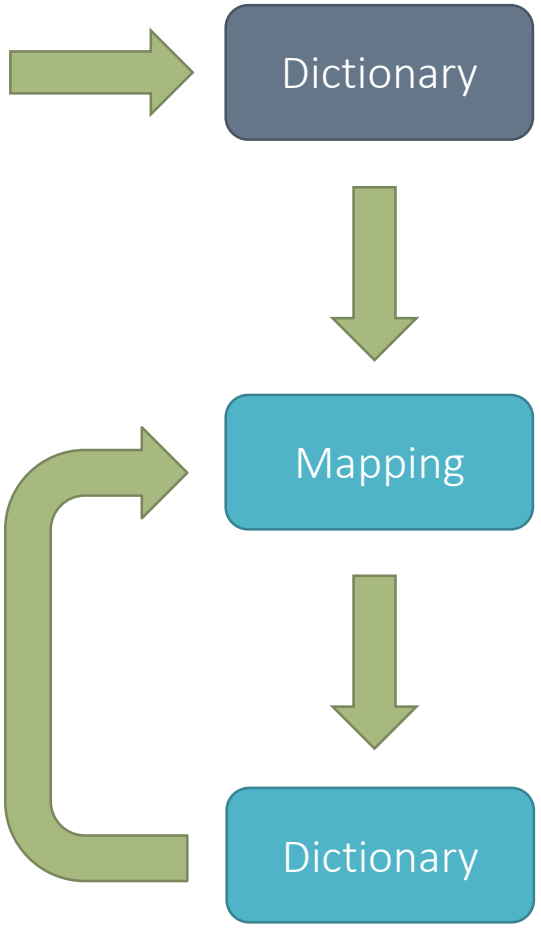


Intra-lingual similarity distribution

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

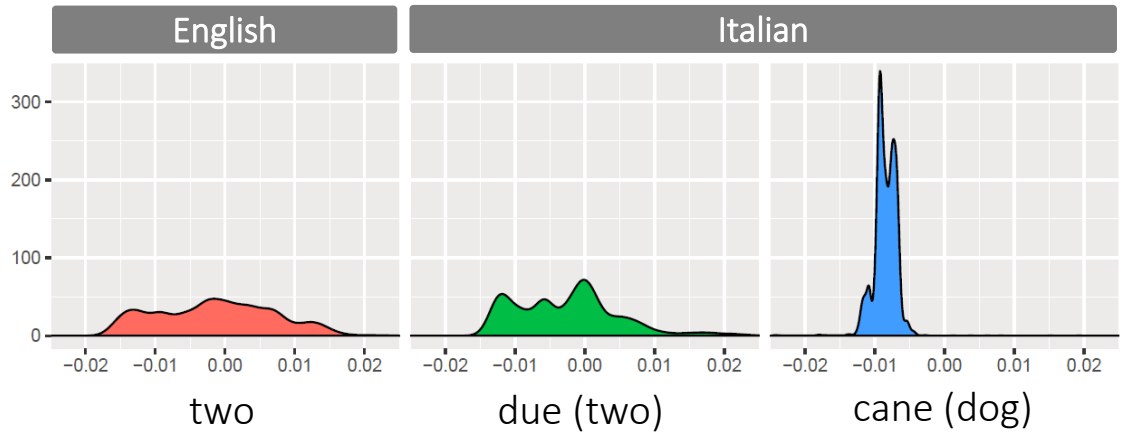
2) Robust self-learning

- Stochastic dictionary induction
- Frequency-based vocabulary cutoff
- CSLS retrieval (Conneau et al., 2018)
- Bidirectional dictionary induction



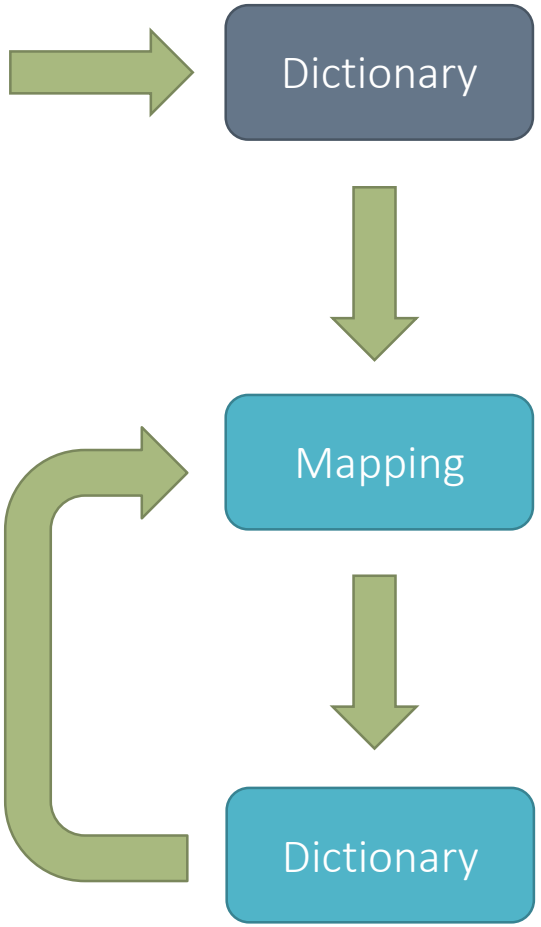
Proposed method

1) Fully unsupervised initialization



Intra-lingual similarity distribution

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$



2) Robust self-learning

- Stochastic dictionary induction
- Frequency-based vocabulary cutoff
- CSLS retrieval (Conneau et al., 2018)
- Bidirectional dictionary induction
- Final symmetric re-weighting (Artetxe et al., 2018)

Experiments

Experiments

- Bilingual lexicon extraction

Experiments

- Bilingual lexicon extraction

Method

Experiments

- Bilingual lexicon extraction

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$

Experiments

- Bilingual lexicon extraction

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper

Experiments

- Bilingual lexicon extraction

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method

  ⇒ *Best/average accuracy*

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$			
Zhang et al. (2017), $\lambda = 10$			
Conneau et al. (2018), code			
Conneau et al. (2018), paper			
Proposed method			

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	71.43	60.38	0.00
Zhang et al. (2017), $\lambda = 10$	70.24	57.64	21.07
Conneau et al. (2018), code	76.18	**67.32**	32.64
Conneau et al. (2018), paper	76.15	67.21	29.79
Proposed method	**76.43**	66.96	**36.10**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	71.43	60.38	0.00
Zhang et al. (2017), $\lambda = 10$	70.24	57.64	21.07
Conneau et al. (2018), code	76.18	**67.32**	32.64
Conneau et al. (2018), paper	76.15	67.21	29.79
Proposed method	**76.43**	66.96	**36.10**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	71.43	60.38	0.00
Zhang et al. (2017), $\lambda = 10$	70.24	57.64	21.07
Conneau et al. (2018), code	76.18	**67.32**	32.64
Conneau et al. (2018), paper	76.15	67.21	29.79
Proposed method	**76.43**	66.96	**36.10**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	71.43	60.38	0.00
Zhang et al. (2017), $\lambda = 10$	70.24	57.64	21.07
Conneau et al. (2018), code	76.18	**67.32**	32.64
Conneau et al. (2018), paper	76.15	67.21	29.79
Proposed method	**76.43**	66.96	**36.10**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	68.18	56.45	0.00
Zhang et al. (2017), $\lambda = 10$	66.37	52.60	17.95
Conneau et al. (2018), code	75.82	**67.00**	14.34
Conneau et al. (2018), paper	75.81	60.22	16.48
Proposed method	**76.28**	66.92	**35.93**

Average accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)

Method	ES-EN	IT-EN	TR-EN
Zhang et al. (2017), $\lambda = 1$	10	10	0
Zhang et al. (2017), $\lambda = 10$	10	10	10
Conneau et al. (2018), code	10	10	5
Conneau et al. (2018), paper	10	9	7
Proposed method	10	10	10

Number of successful runs

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method
Zhang et al. (2017), $\lambda = 1$
Zhang et al. (2017), $\lambda = 10$
Conneau et al. (2018), code
Conneau et al. (2018), paper
Proposed method

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$				
Zhang et al. (2017), $\lambda = 10$				
Conneau et al. (2018), code				
Conneau et al. (2018), paper				
Proposed method				

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0.00	0.00	0.00	0.00
Zhang et al. (2017), $\lambda = 10$	0.00	0.00	0.01	0.01
Conneau et al. (2018), code	45.40	47.27	1.62	36.20
Conneau et al. (2018), paper	45.27	0.07	0.07	35.47
Proposed method	**48.53**	**48.47**	**33.50**	**37.60**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0.00	0.00	0.00	0.00
Zhang et al. (2017), $\lambda = 10$	0.00	0.00	0.01	0.01
Conneau et al. (2018), code	45.40	47.27	1.62	36.20
Conneau et al. (2018), paper	45.27	0.07	0.07	35.47
Proposed method	**48.53**	**48.47**	**33.50**	**37.60**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0.00	0.00	0.00	0.00
Zhang et al. (2017), $\lambda = 10$	0.00	0.00	0.01	0.01
Conneau et al. (2018), code	45.40	47.27	1.62	36.20
Conneau et al. (2018), paper	45.27	0.07	0.07	35.47
Proposed method	**48.53**	**48.47**	**33.50**	**37.60**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0.00	0.00	0.00	0.00
Zhang et al. (2017), $\lambda = 10$	0.00	0.00	0.01	0.01
Conneau et al. (2018), code	45.40	47.27	1.62	36.20
Conneau et al. (2018), paper	45.27	0.07	0.07	35.47
Proposed method	**48.53**	**48.47**	**33.50**	**37.60**

Best accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0.00	0.00	0.00	0.00
Zhang et al. (2017), $\lambda = 10$	0.00	0.00	0.01	0.01
Conneau et al. (2018), code	13.55	42.15	0.38	21.23
Conneau et al. (2018), paper	9.10	0.01	0.01	7.09
Proposed method	**48.13**	**48.19**	**32.63**	**37.33**

Average accuracy (%)

Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  - ⇒ *Best/average accuracy*
  - ⇒ *Successful runs (>5% accuracy)*
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

Method	EN-IT	EN-DE	EN-FI	EN-ES
Zhang et al. (2017), $\lambda = 1$	0	0	0	0
Zhang et al. (2017), $\lambda = 10$	0	0	0	0
Conneau et al. (2018), code	3	9	0	6
Conneau et al. (2018), paper	2	0	0	2
Proposed method	10	10	10	10

Number of successful runs

Experiments

Supervision	Method

Experiments

Supervision	Method
	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
5k dict.	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Artetxe et al. (2017)
	Smith et al. (2017)
	Artetxe et al. (2018)

Experiments

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Artetxe et al. (2017)
	Smith et al. (2017)
	Artetxe et al. (2018)
25 dict.	Artetxe et al. (2017)
Init. heurist.	Smith et al. (2017), cognates
	Artetxe et al. (2017), num.

Experiments

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Artetxe et al. (2017)
	Smith et al. (2017)
	Artetxe et al. (2018)
25 dict.	Artetxe et al. (2017)
Init. heurist.	Smith et al. (2017), cognates
	Artetxe et al. (2017), num.
None	Zhang et al. (2017), $\lambda = 1$
	Zhang et al. (2017), $\lambda = 10$
	Conneau et al. (2018), code‡
	Conneau et al. (2018), paper‡
	Proposed method

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)				
	Faruqui and Dyer (2014)				
	Shigeto et al. (2015)				
	Dinu et al. (2015)				
	Lazaridou et al. (2015)				
	Xing et al. (2015)				
	Zhang et al. (2016)				
	Artetxe et al. (2016)				
	Artetxe et al. (2017)				
	Smith et al. (2017)				
	Artetxe et al. (2018)				
25 dict.	Artetxe et al. (2017)				
Init. heurist.	Smith et al. (2017), cognates				
	Artetxe et al. (2017), num.				
None	Zhang et al. (2017), $\lambda = 1$				
	Zhang et al. (2017), $\lambda = 10$				
	Conneau et al. (2018), code‡				
	Conneau et al. (2018), paper‡				
	Proposed method				

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

Experiments

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Artetxe et al. (2017)	39.67	40.87	28.72	-
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018)	45.27	44.13	**32.94**	36.60
25 dict.	Artetxe et al. (2017)	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Proposed method	**48.13**	**48.19**	32.63	**37.33**

Conclusions

Conclusions

- Not a solved problem!

Conclusions

- Not a solved problem!
- New self-learning approach

Conclusions

- Not a solved problem!

- New self-learning approach

  ⇒ *Fully unsupervised initialization*

Conclusions

- Not a solved problem!

- New self-learning approach
  - ⇒ *Fully unsupervised initialization*
  - ⇒ *Robust self-learning*

Conclusions

- Not a solved problem!

- New self-learning approach
  - ⇒ *Fully unsupervised initialization*
  - ⇒ *Robust self-learning*

- More robust and accurate than previous methods

Conclusions

- Not a solved problem!
- New self-learning approach
  - ⇒ *Fully unsupervised initialization*
  - ⇒ *Robust self-learning*
- More robust and accurate than previous methods
- Future work: from bilingual to multilingual

One more thing…

>

One more thing…

```
> git clone https://github.com/artetxem/vecmap.git
>
```

One more thing…

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py
```

One more thing…

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --unsupervised
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --unsupervised \
  SRC.EMB TRG.EMB
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --unsupervised \
  SRC.EMB TRG.EMB SRC_MAPPED.EMB TRG_MAPPED.EMB
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --unsupervised \
  SRC.EMB TRG.EMB SRC_MAPPED.EMB TRG_MAPPED.EMB
>
```

One more thing…

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --unsupervised \
  SRC.EMB TRG.EMB SRC_MAPPED.EMB TRG_MAPPED.EMB
>
```

https://github.com/artetxem/vecmap

Thank you!

https://github.com/artetxem/vecmap