

Learning bilingual word embeddings with (almost) no bilingual data

Mikel Artetxe, Gorka Labaka, Eneko Agirre

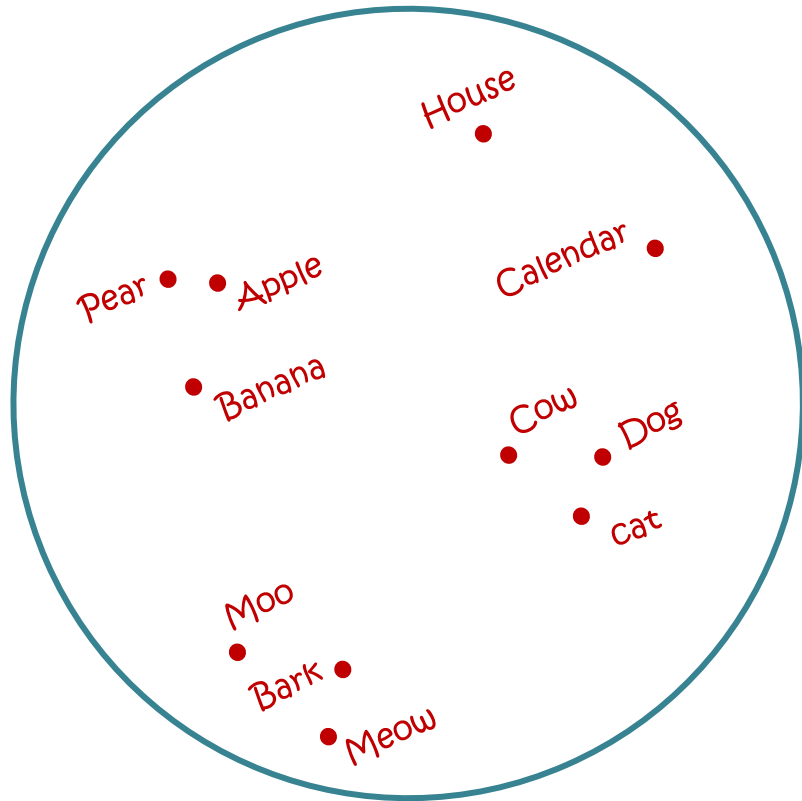
IXA NLP group – University of the Basque Country (UPV/EHU)

Who cares?

Who cares?

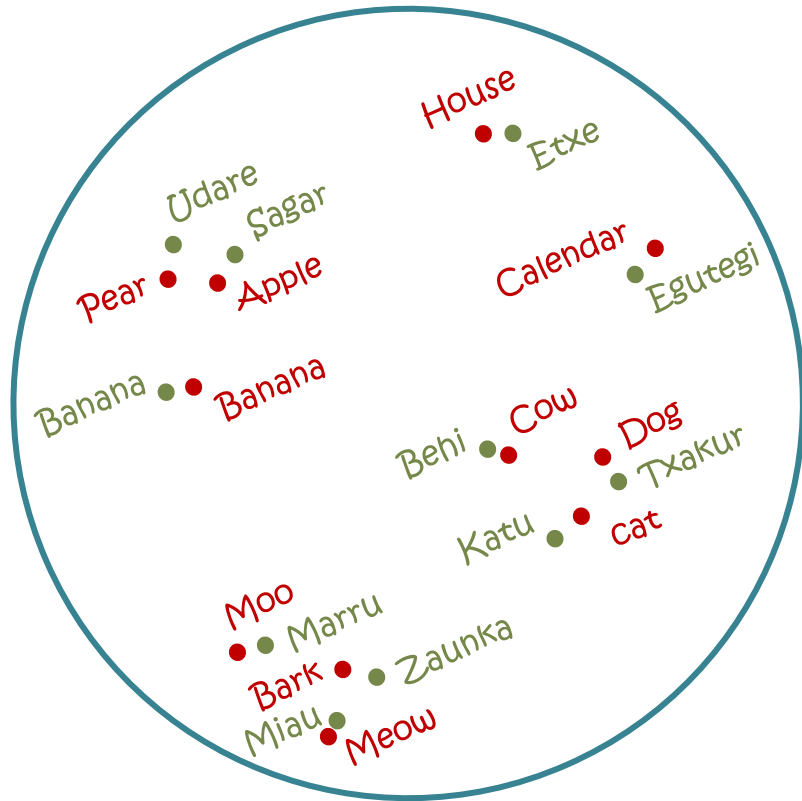
word embeddings are useful!

Who cares?



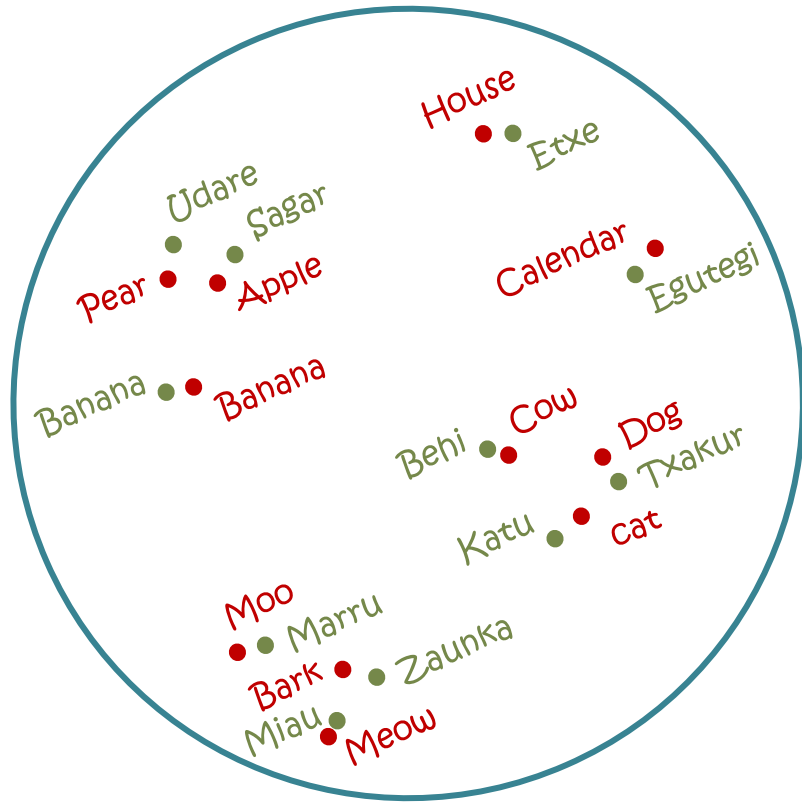
word embeddings are useful!

Who cares?



word embeddings are useful!

Who cares?

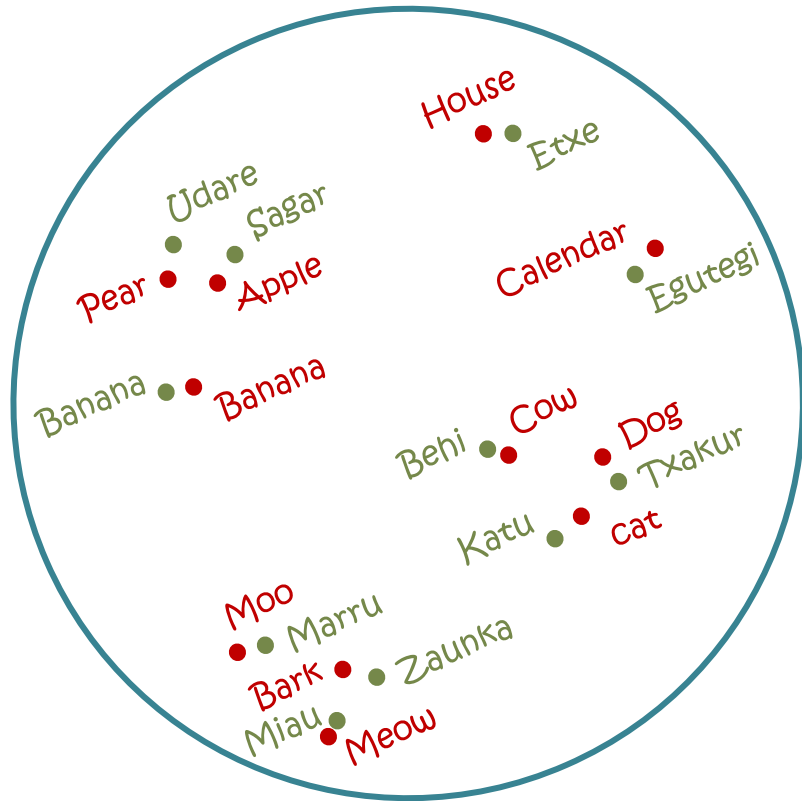


bilingual

word embeddings are useful!

even more

Who cares?



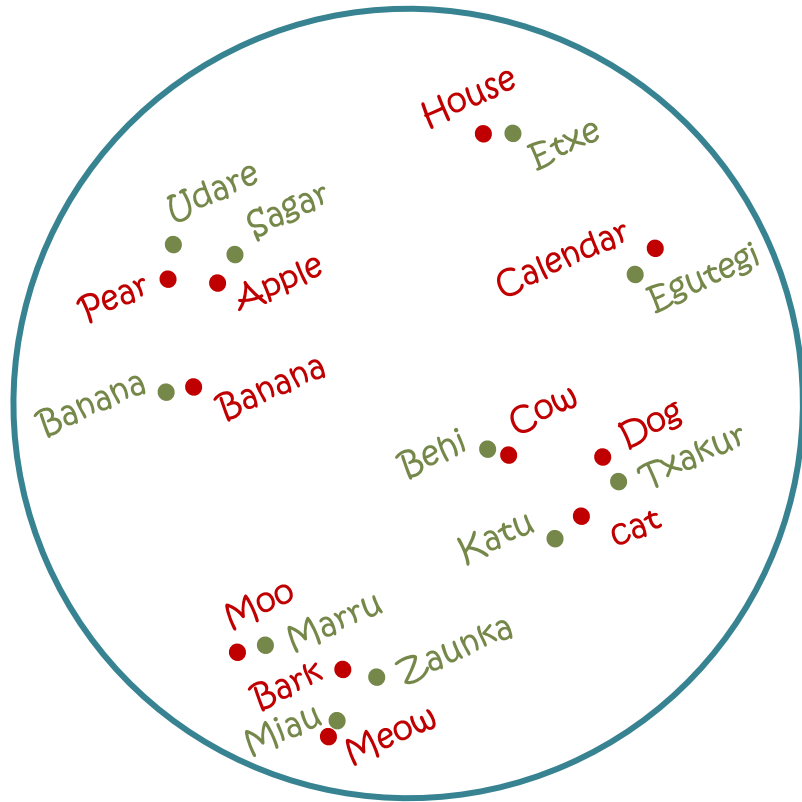
bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks

Who cares?



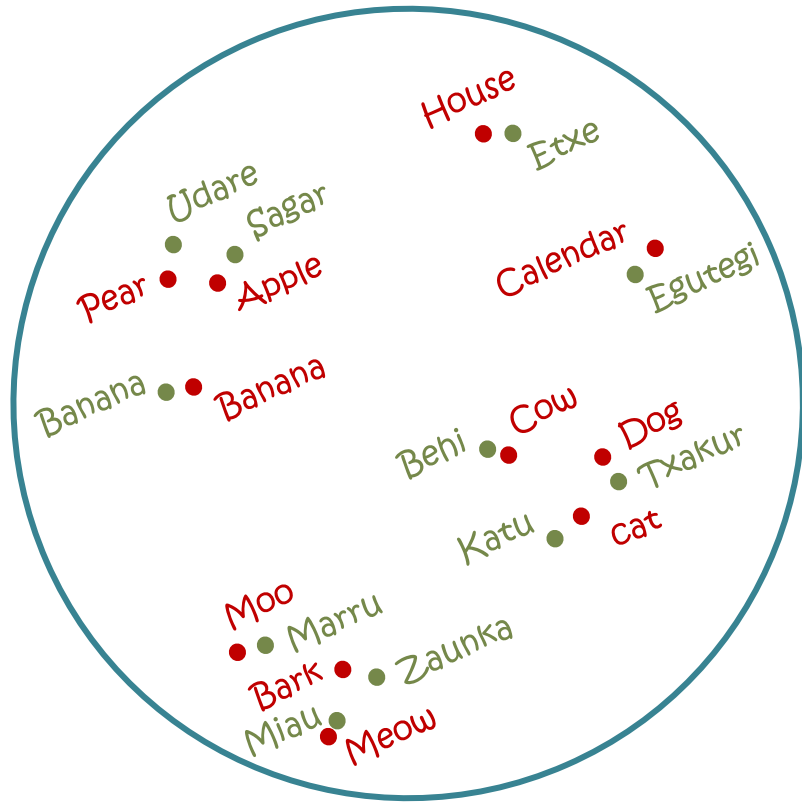
bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks
- crosslingual transfer learning

Who cares?



bilingual signal
for training

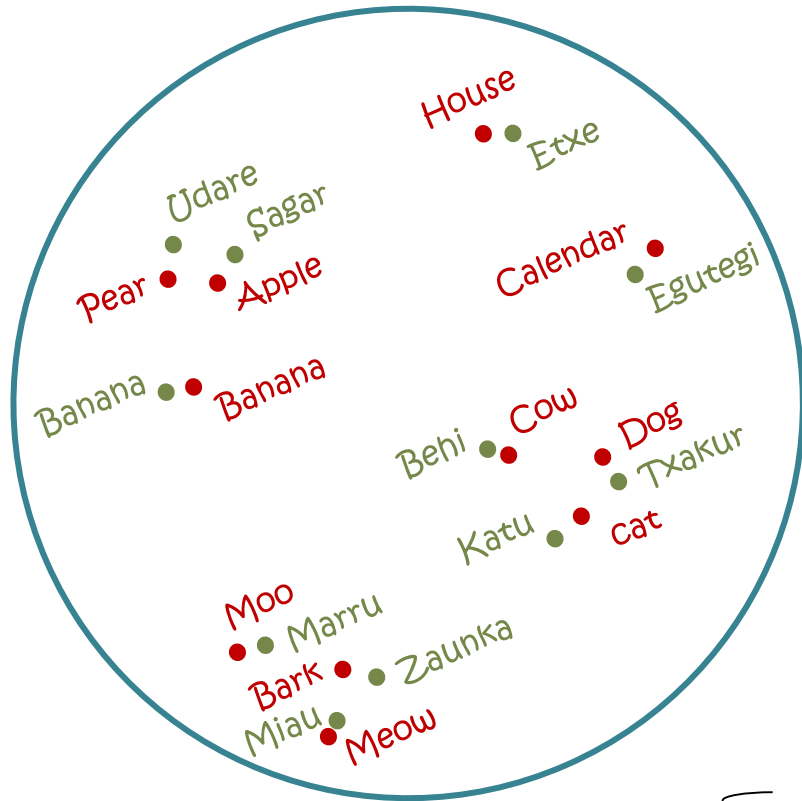
bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks
- crosslingual transfer learning

Who cares?



bilingual signal
for training

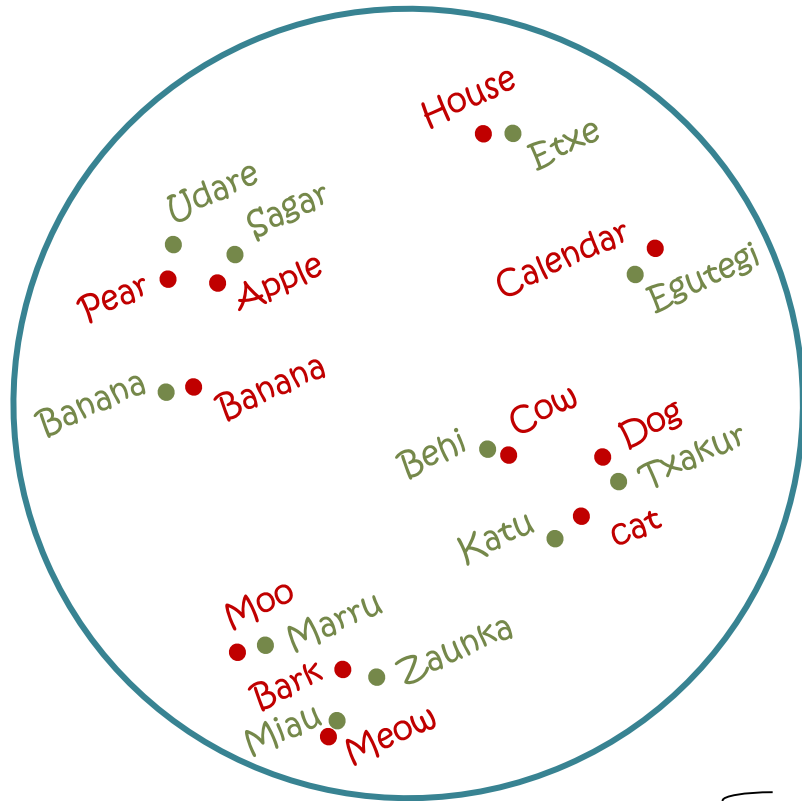
bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks
- crosslingual transfer learning

Who cares?



bilingual

word embeddings are useful!

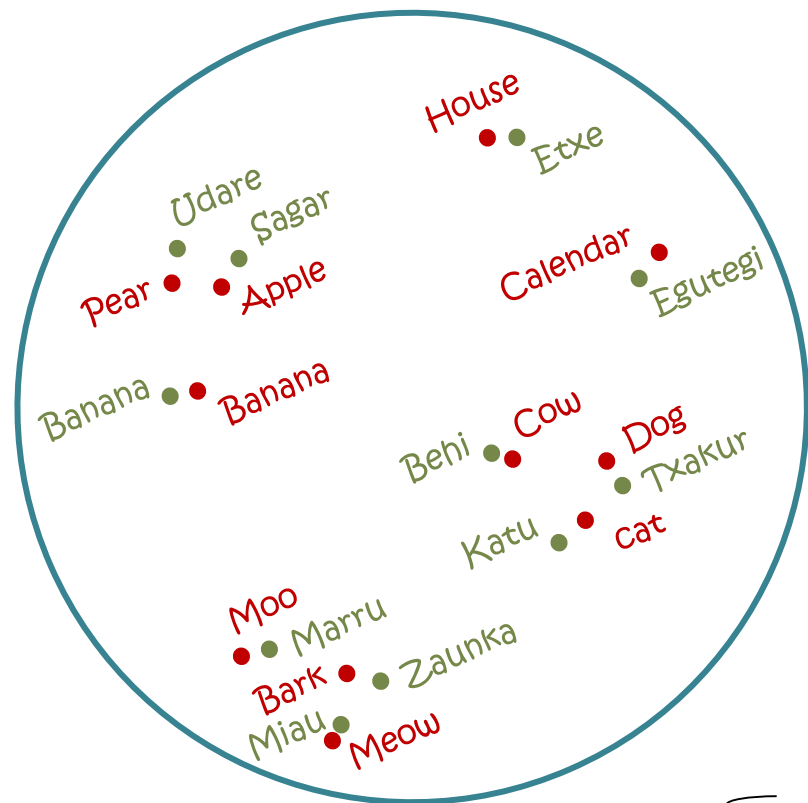
even more

- inherently crosslingual tasks
- crosslingual transfer learning

Previous work

bilingual signal
for training

Who cares?



bilingual signal
for training

bilingual

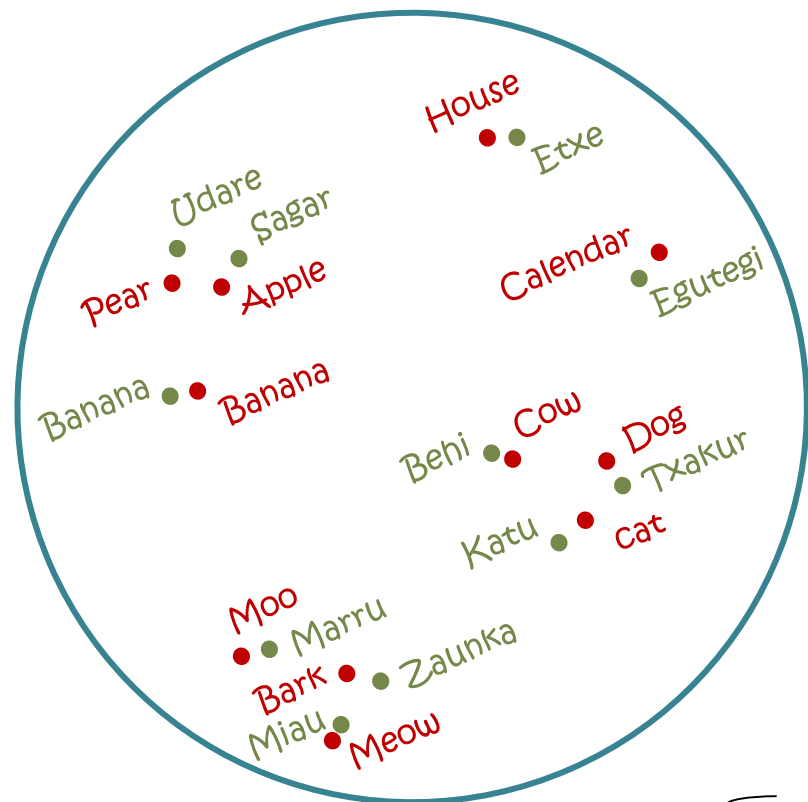
word embeddings are useful!
even more

- inherently crosslingual tasks
- crosslingual transfer learning

Previous work

- parallel corpora

Who cares?



bilingual signal
for training

bilingual

word embeddings are useful!

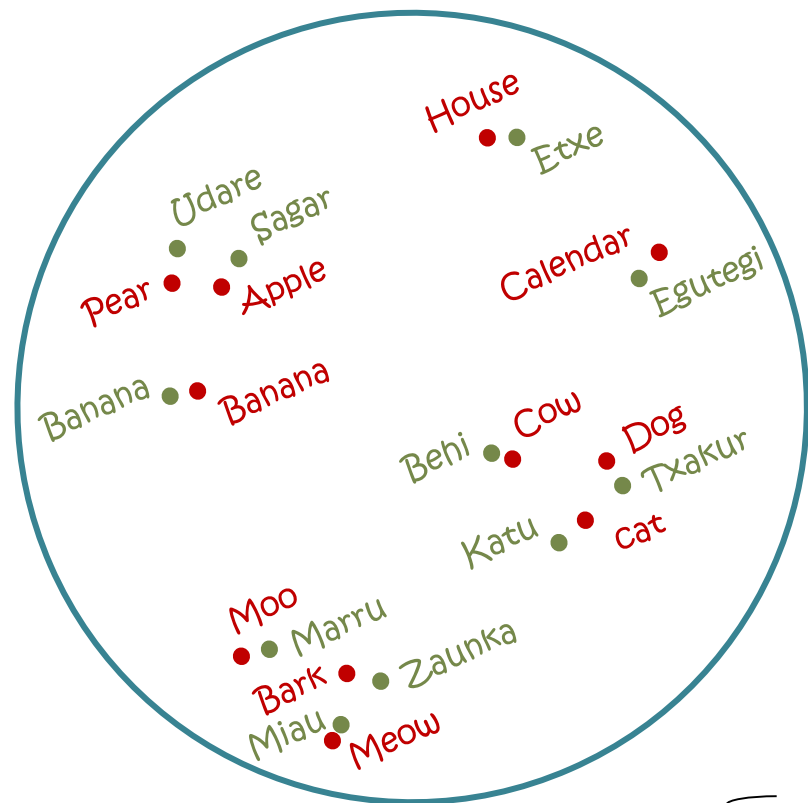
even more

- inherently crosslingual tasks
- crosslingual transfer learning

Previous work

- parallel corpora
- comparable corpora

Who cares?



bilingual signal
for training

bilingual

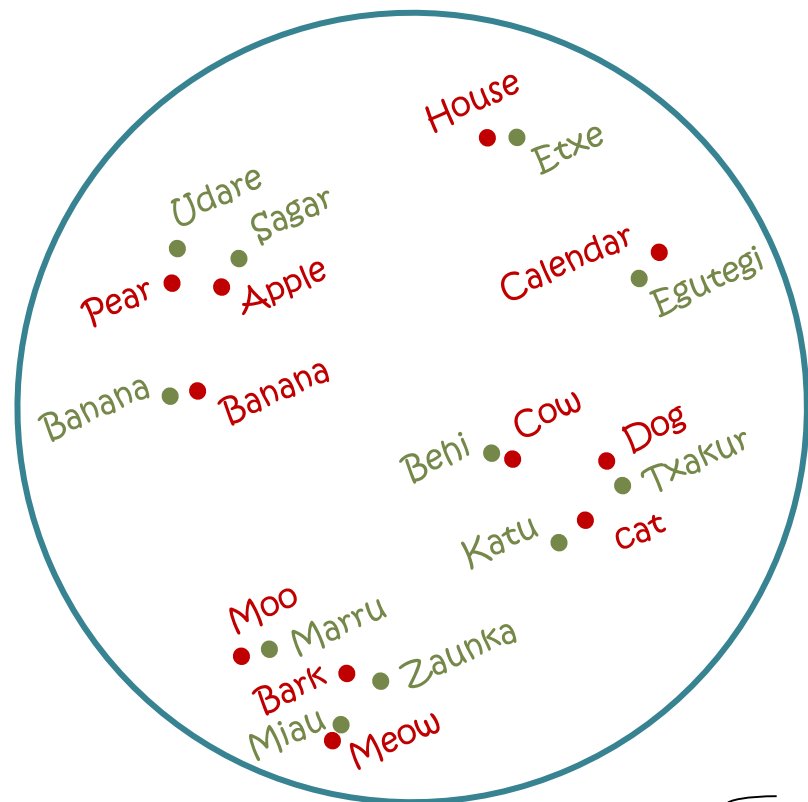
word embeddings are useful!
even more

- inherently crosslingual tasks
- crosslingual transfer learning

Previous work

- parallel corpora
- comparable corpora
- (big) dictionaries

Who cares?



bilingual signal
for training

bilingual

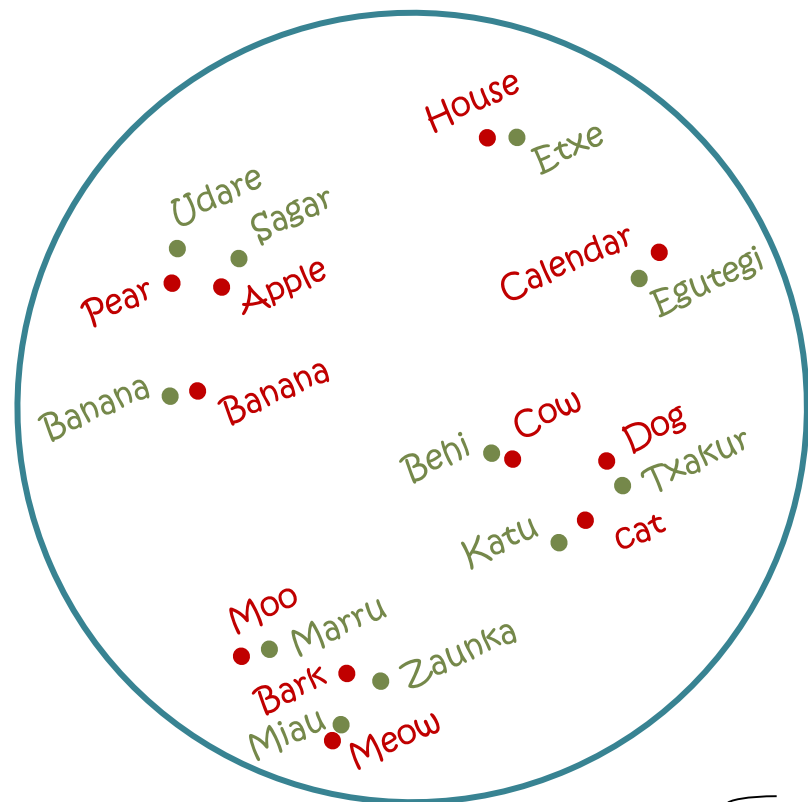
word embeddings are useful!
even more

- inherently crosslingual tasks
- crosslingual transfer learning

~~Previous work~~

- ~~- parallel corpora~~
- ~~- comparable corpora~~
- ~~- (big) dictionaries~~

Who cares?



bilingual signal
for training

bilingual

word embeddings are useful!

even more

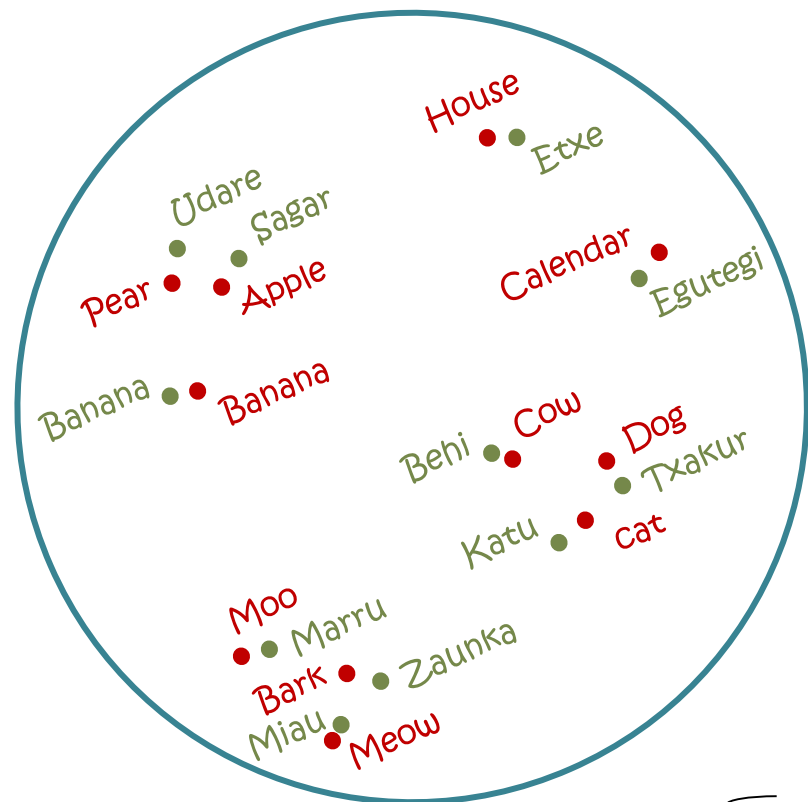
- inherently crosslingual tasks
- crosslingual transfer learning

~~Previous work~~

- ~~- parallel corpora~~
- ~~- comparable corpora~~
- ~~- (bilingual) dictionaries~~

This talk

Who cares?



bilingual signal
for training

bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks
- crosslingual transfer learning

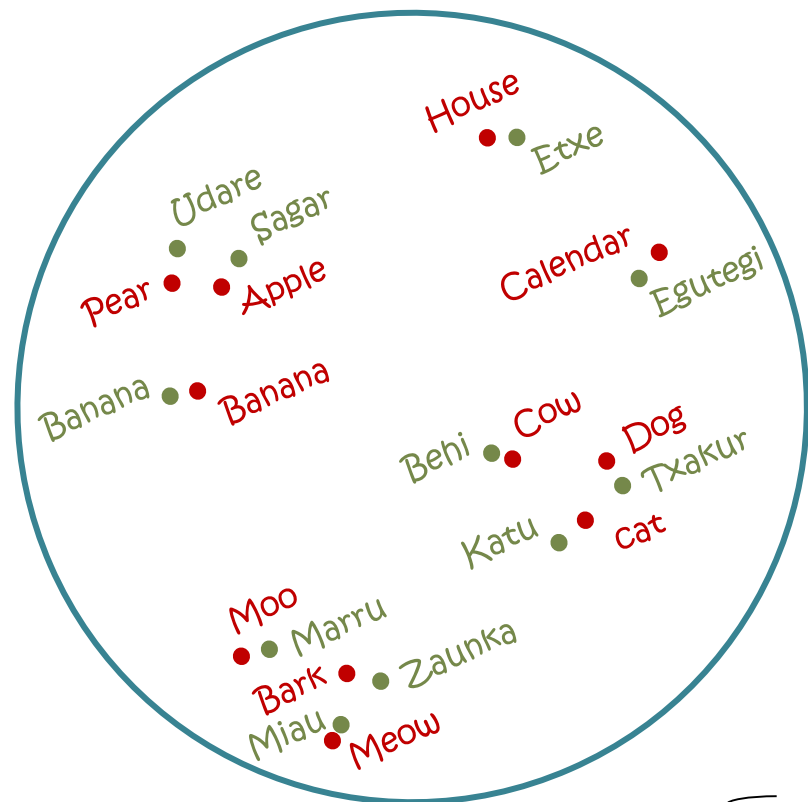
~~Previous work~~

- ~~- parallel corpora~~
- ~~- comparable corpora~~
- ~~- (big) dictionaries~~

This talk

- 25 word dictionary

Who cares?



bilingual signal
for training

bilingual

word embeddings are useful!

even more

- inherently crosslingual tasks
- crosslingual transfer learning

~~Previous work~~

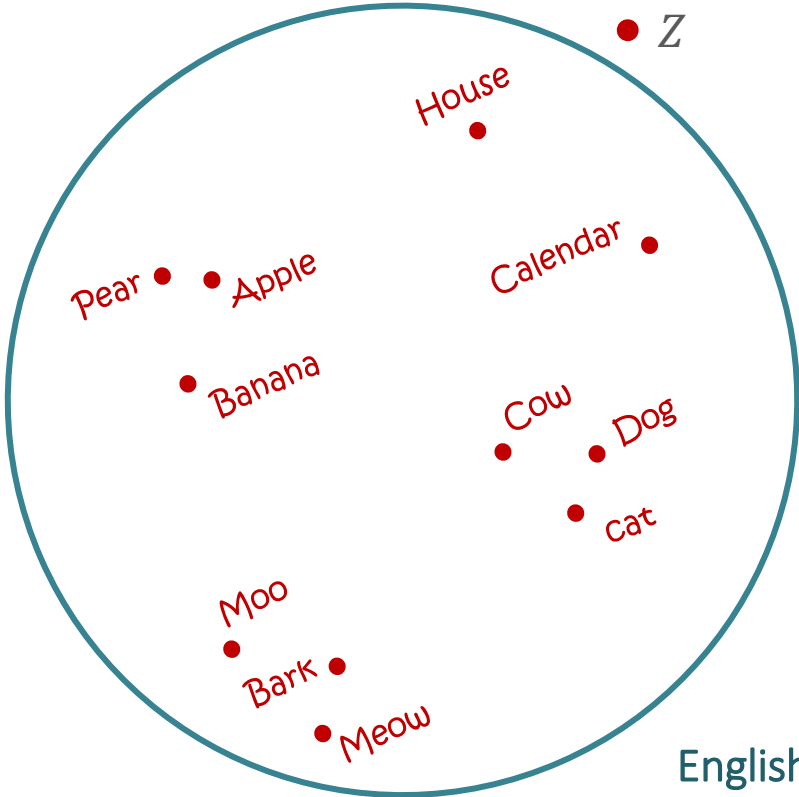
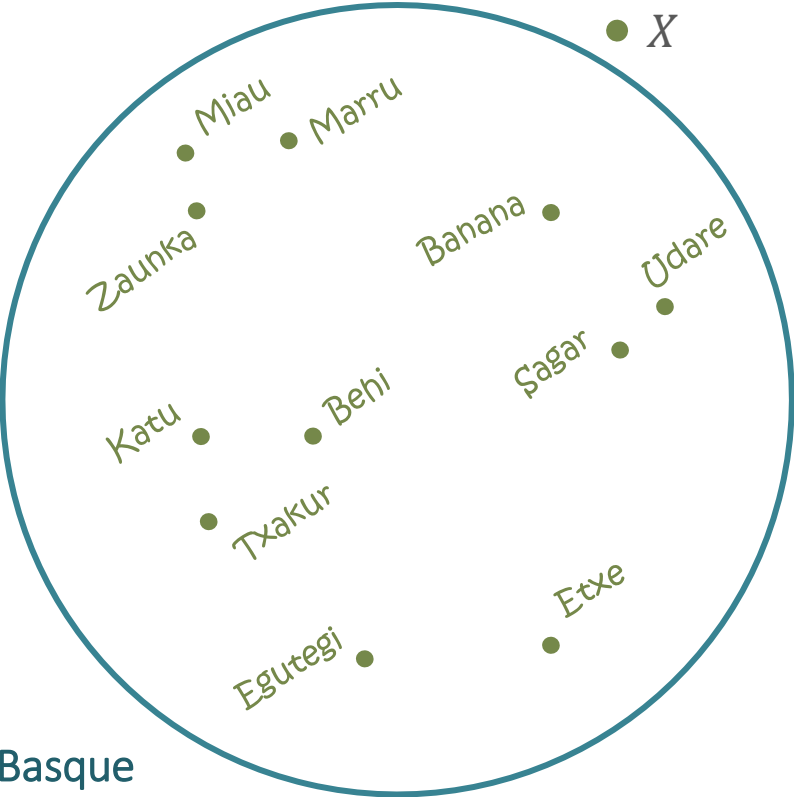
- ~~- parallel corpora~~
- ~~- comparable corpora~~
- ~~- (big) dictionaries~~

This talk

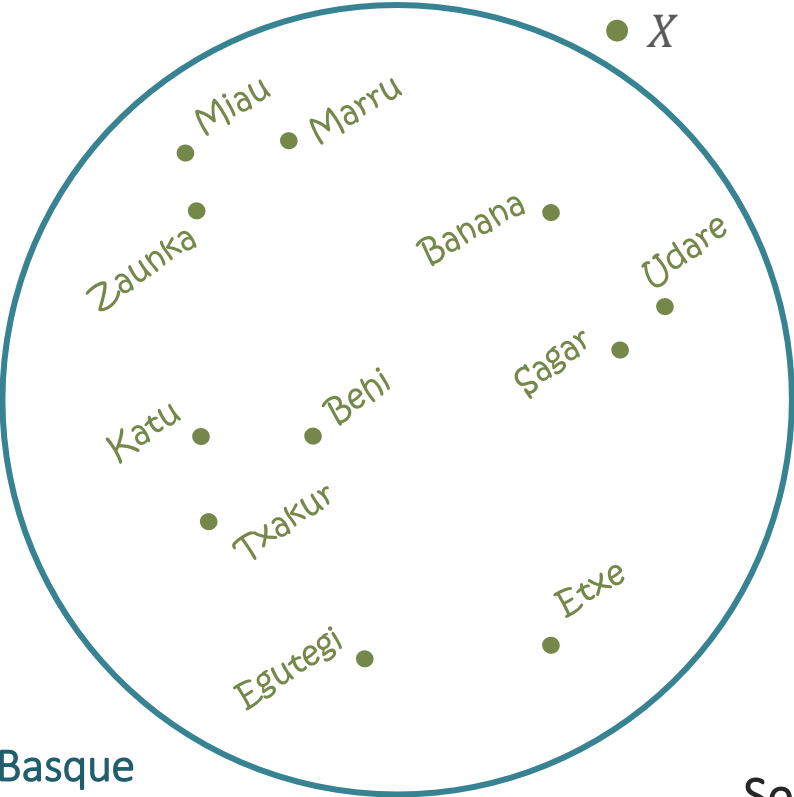
- 25 word dictionary
- numerals (1, 2, 3...)

Bilingual embedding mappings

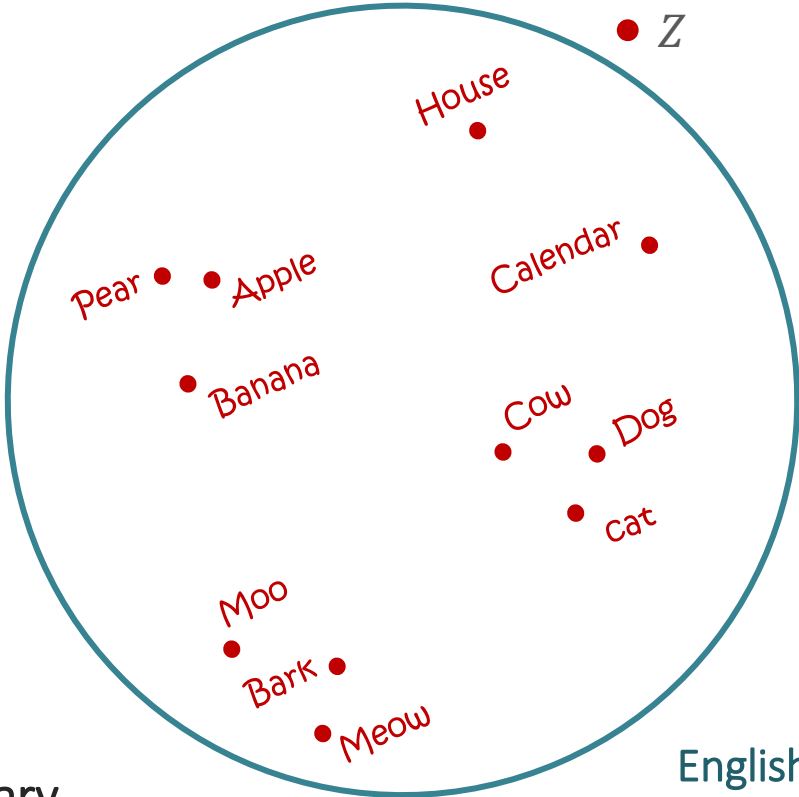
Bilingual embedding mappings



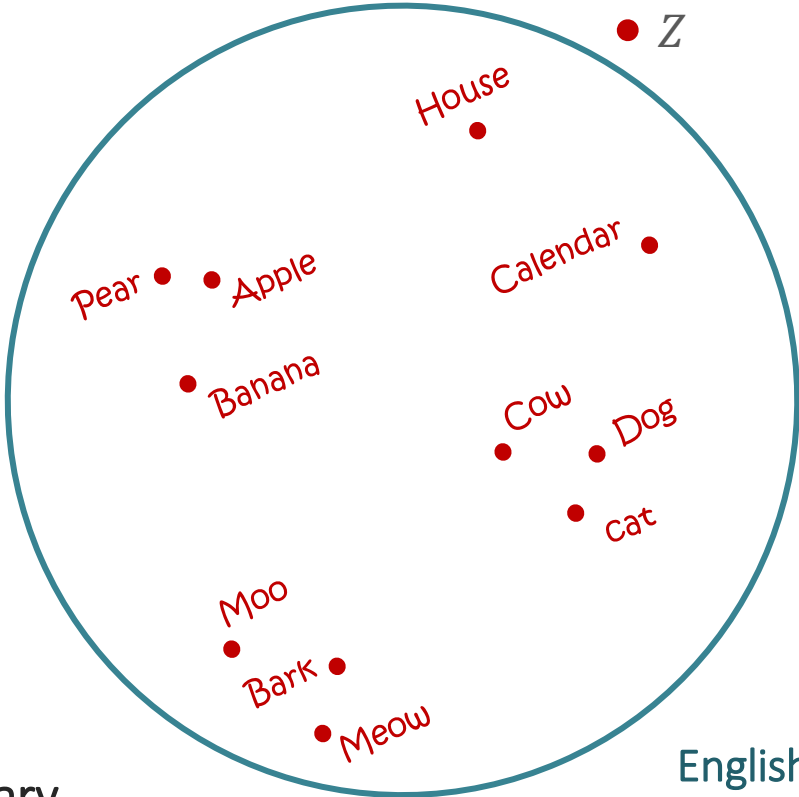
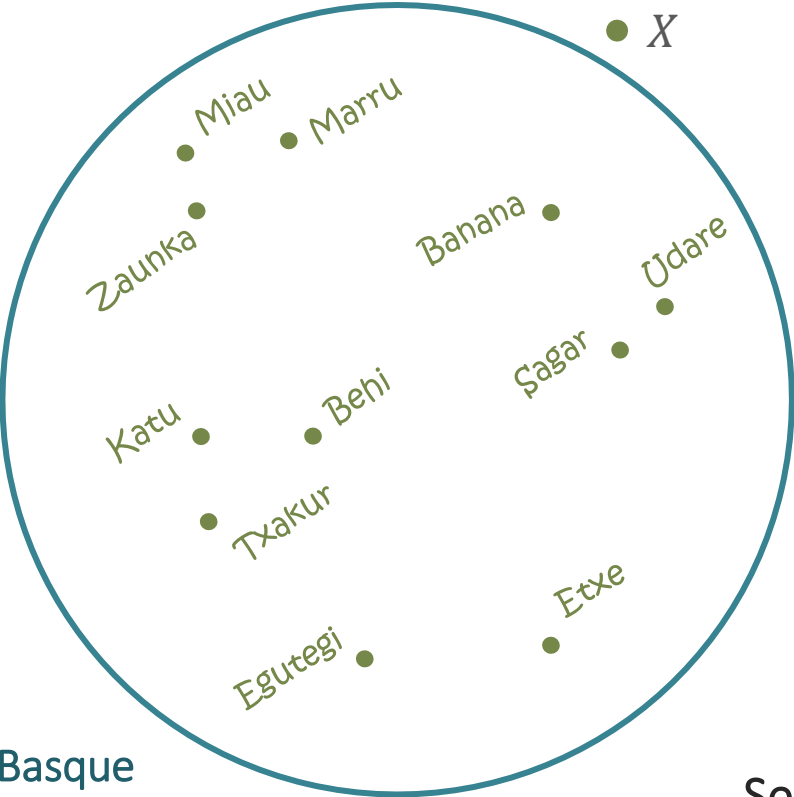
Bilingual embedding mappings



Seed dictionary



Bilingual embedding mappings

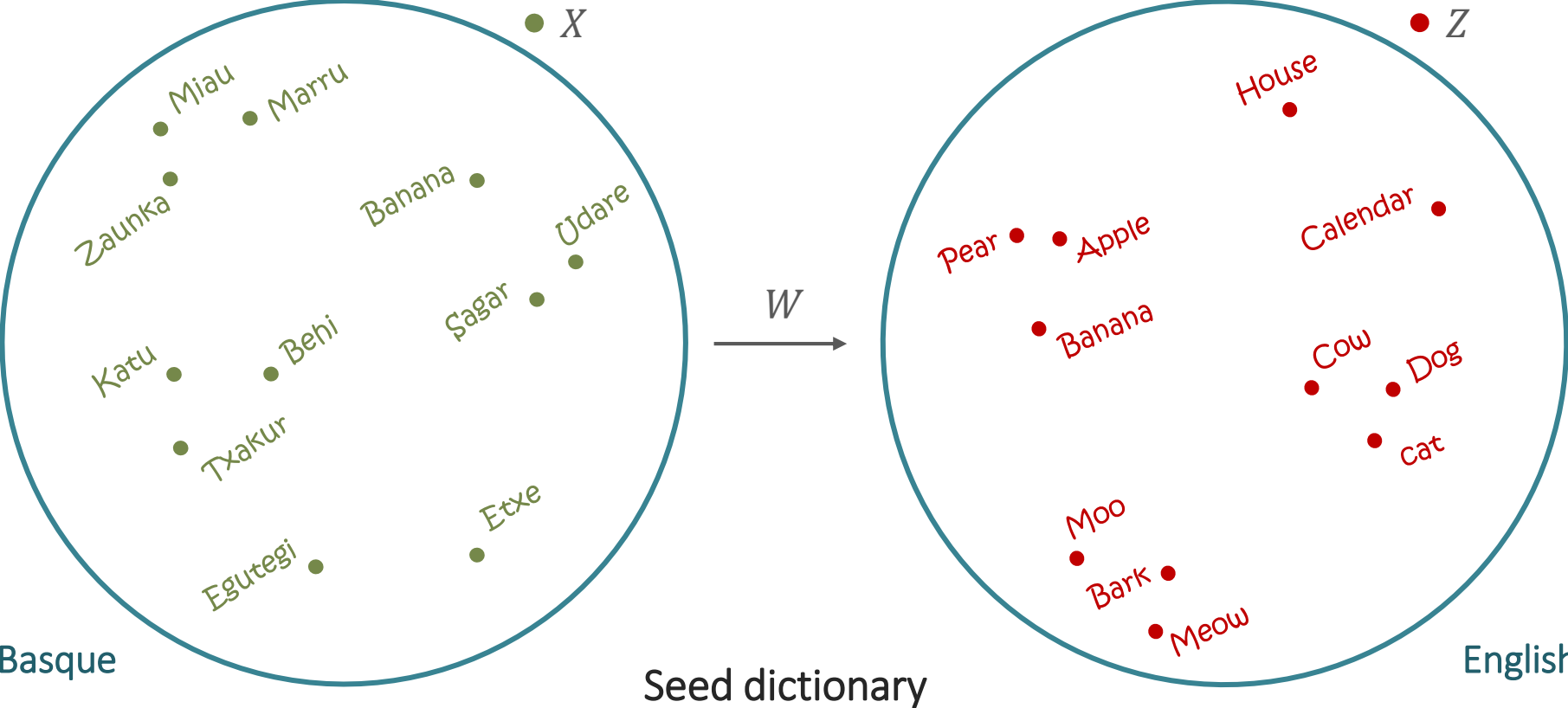


Seed dictionary

Txakur
Sagar
⋮
Egutegi

Dog
Apple
⋮
Calendar

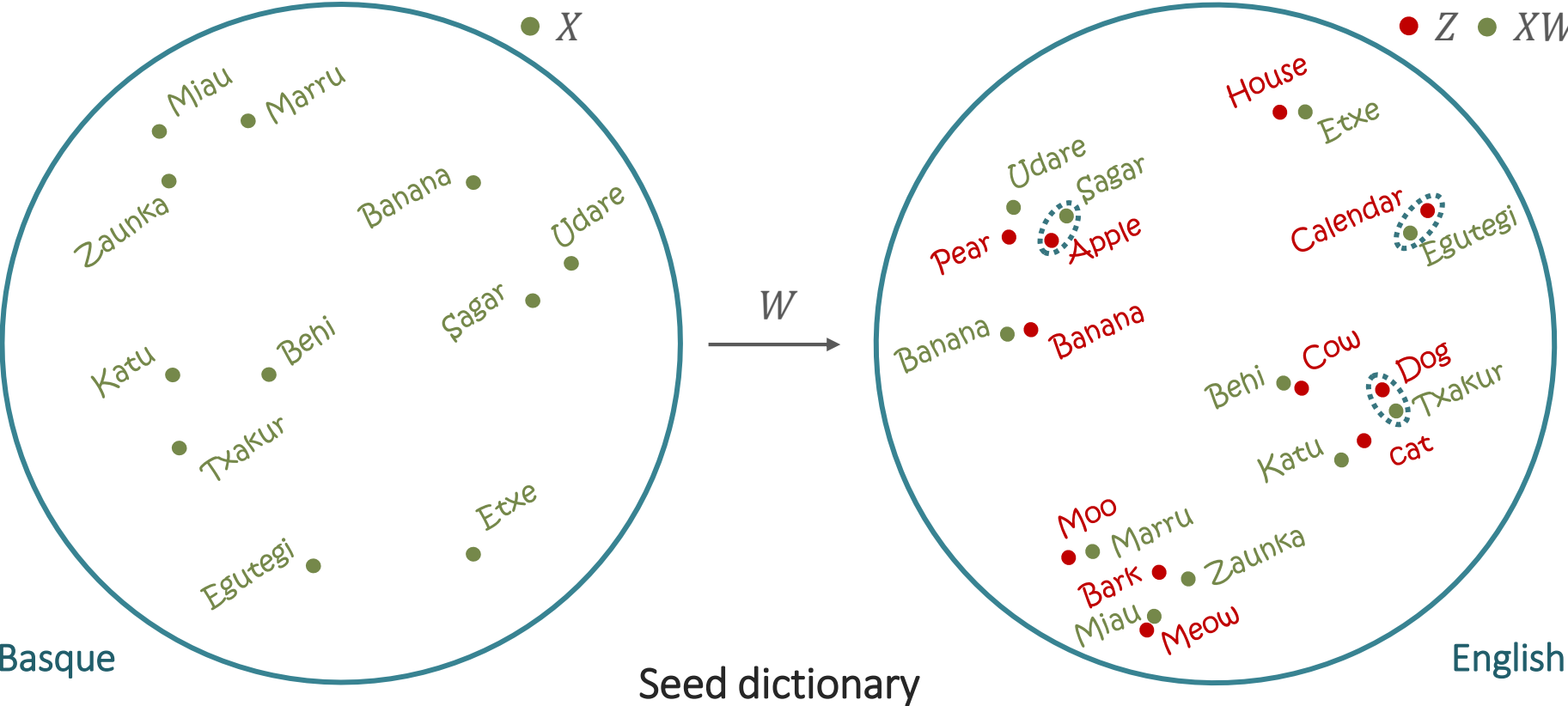
Bilingual embedding mappings



Txakur
Sagar
⋮
Egutegi

Dog
Apple
⋮
Calendar

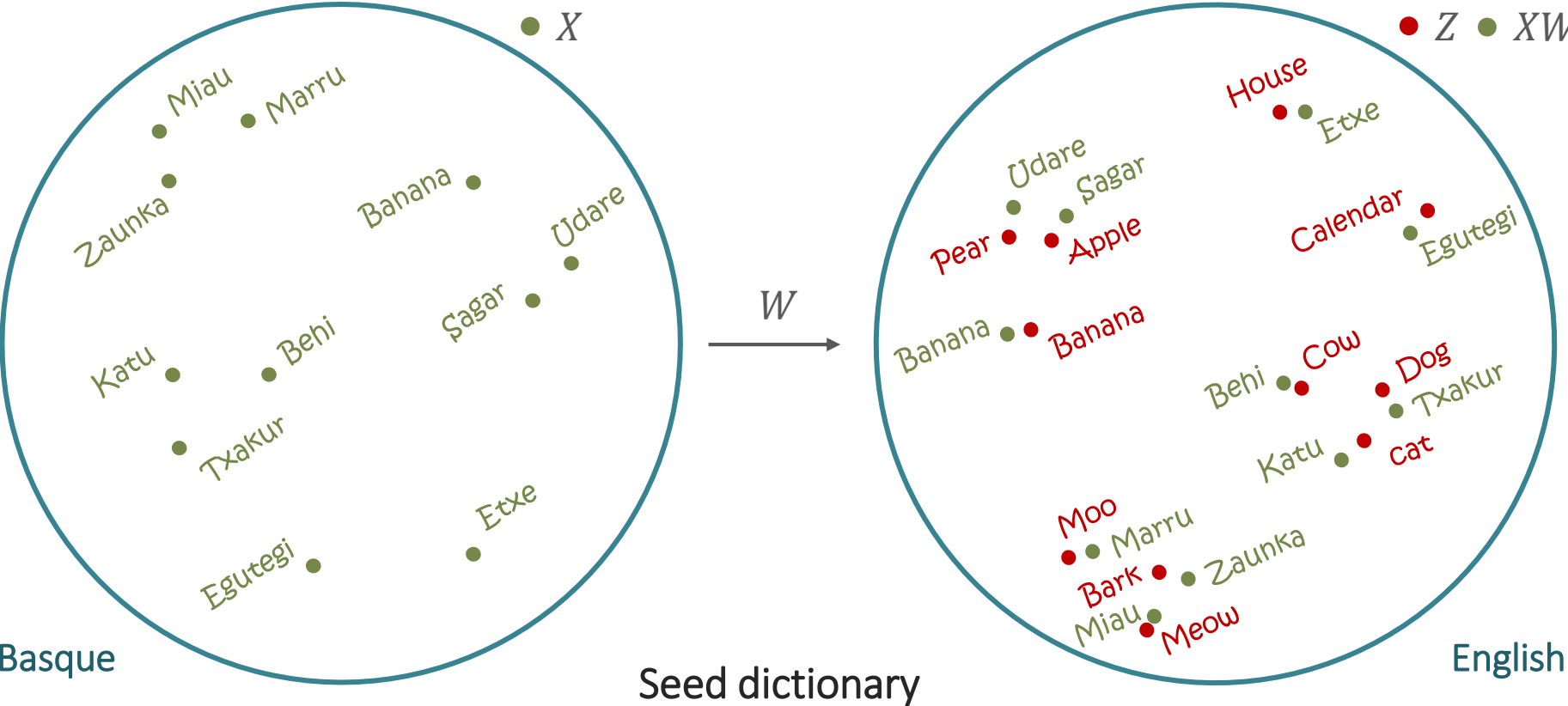
Bilingual embedding mappings



Txakur
Sagar
⋮
Egutegi

Dog
Apple
⋮
Calendar

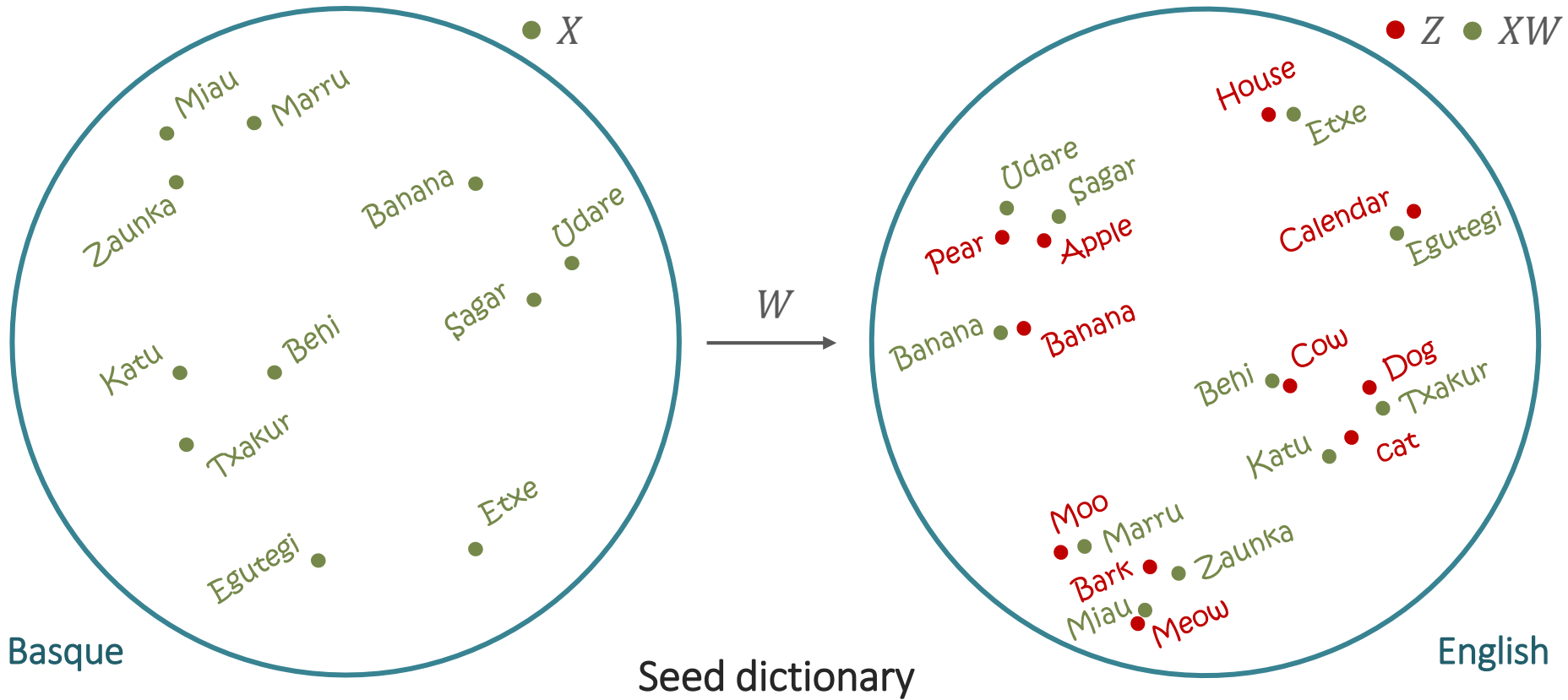
Bilingual embedding mappings



| | |
|---------|---|
| Txakur | $\begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix}$ |
| Sagar | |
| ⋮ | |
| Egutegi | |

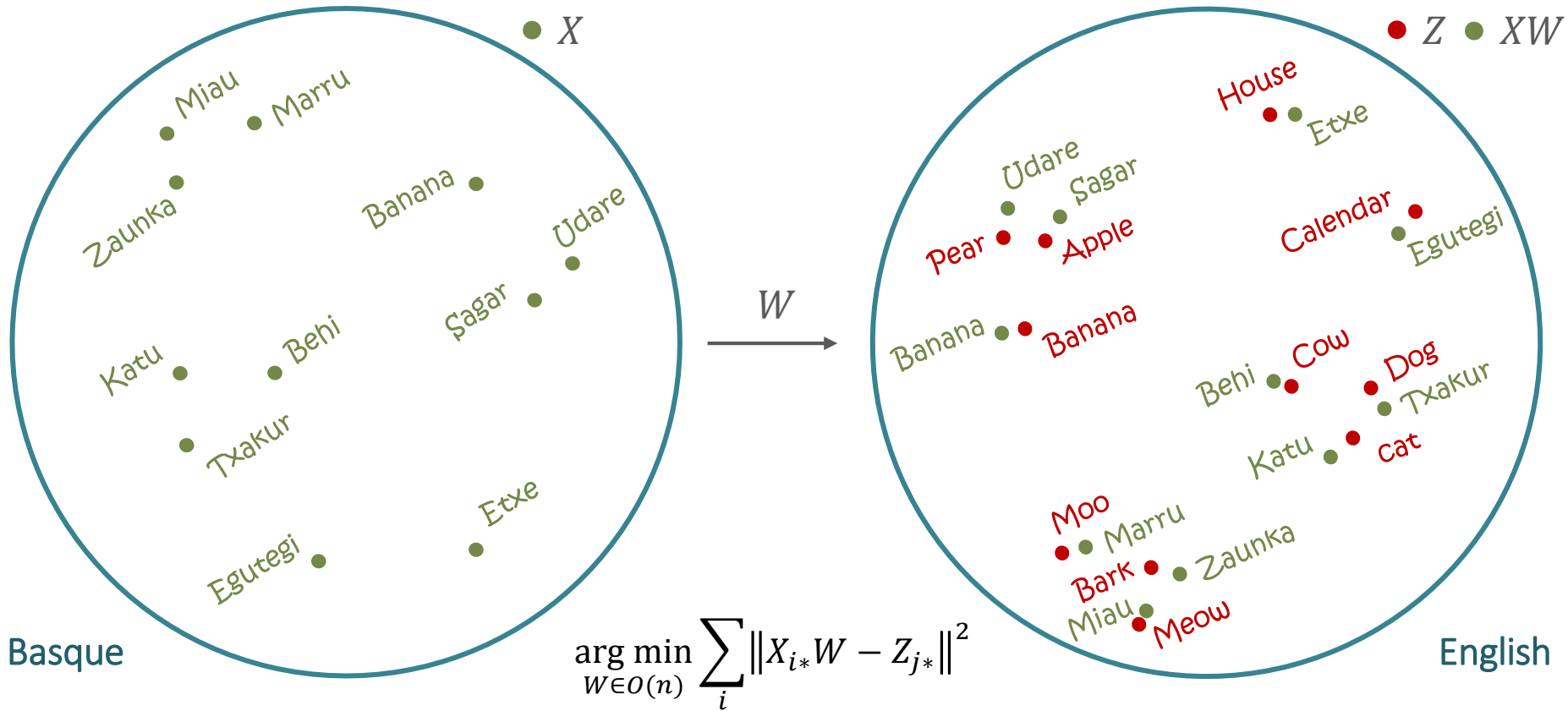
| | |
|---|----------|
| $\begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix}$ | Dog |
| | Apple |
| | ⋮ |
| | Calendar |

Bilingual embedding mappings



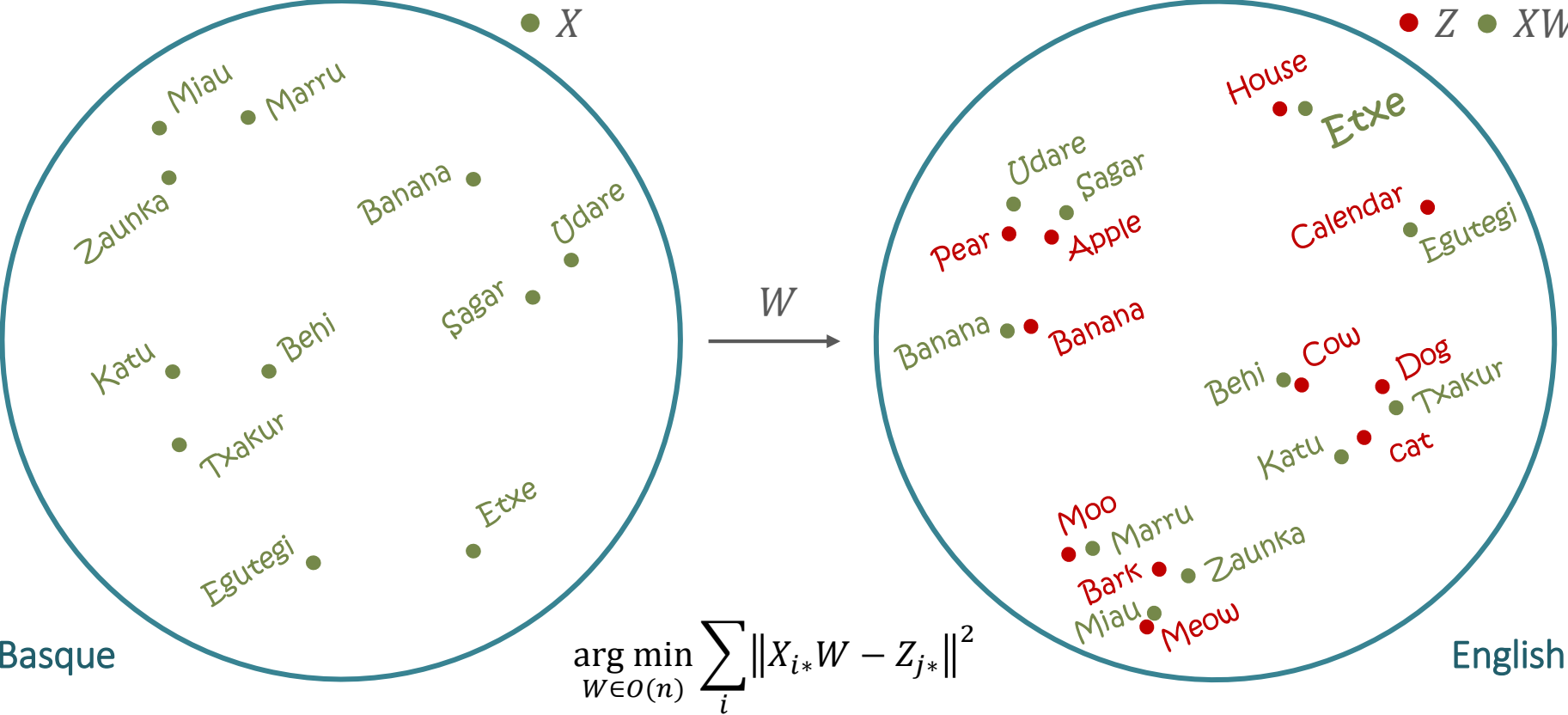
$$\begin{array}{l}
 \text{Txakur} \\
 \text{Sagar} \\
 \vdots \\
 \text{Egutegi}
 \end{array}
 \begin{bmatrix}
 X_{1,*} \\
 X_{2,*} \\
 \vdots \\
 X_{n,*}
 \end{bmatrix}
 [W] \approx
 \begin{bmatrix}
 Z_{1,*} \\
 Z_{2,*} \\
 \vdots \\
 Z_{n,*}
 \end{bmatrix}
 \begin{array}{l}
 \text{Dog} \\
 \text{Apple}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

Bilingual embedding mappings



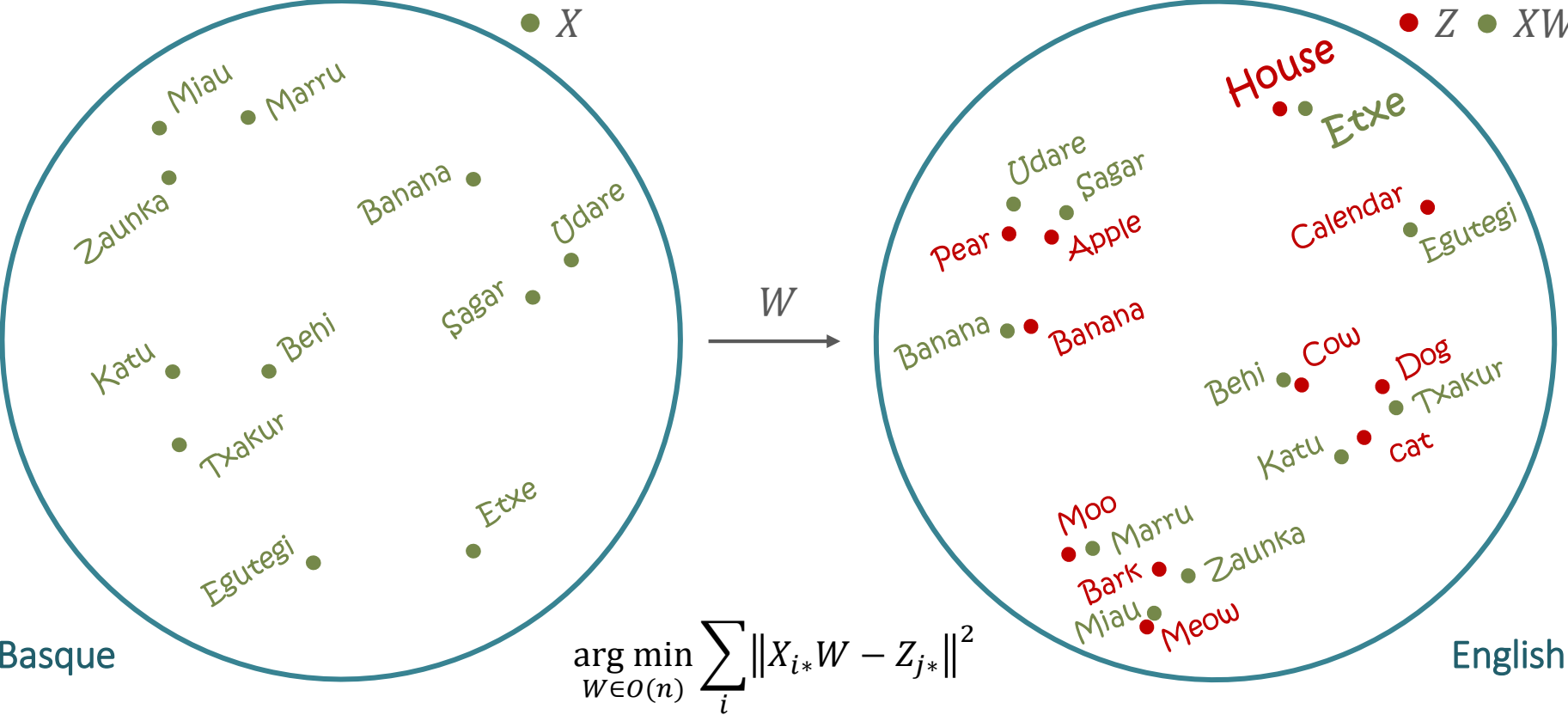
$$\begin{array}{l}
 \text{Txakur} \\
 \text{Sagar} \\
 \vdots \\
 \text{Egutegi}
 \end{array}
 \begin{bmatrix}
 X_{1,*} \\
 X_{2,*} \\
 \vdots \\
 X_{n,*}
 \end{bmatrix}
 [W] \approx
 \begin{bmatrix}
 Z_{1,*} \\
 Z_{2,*} \\
 \vdots \\
 Z_{n,*}
 \end{bmatrix}
 \begin{array}{l}
 \text{Dog} \\
 \text{Apple}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

Bilingual embedding mappings



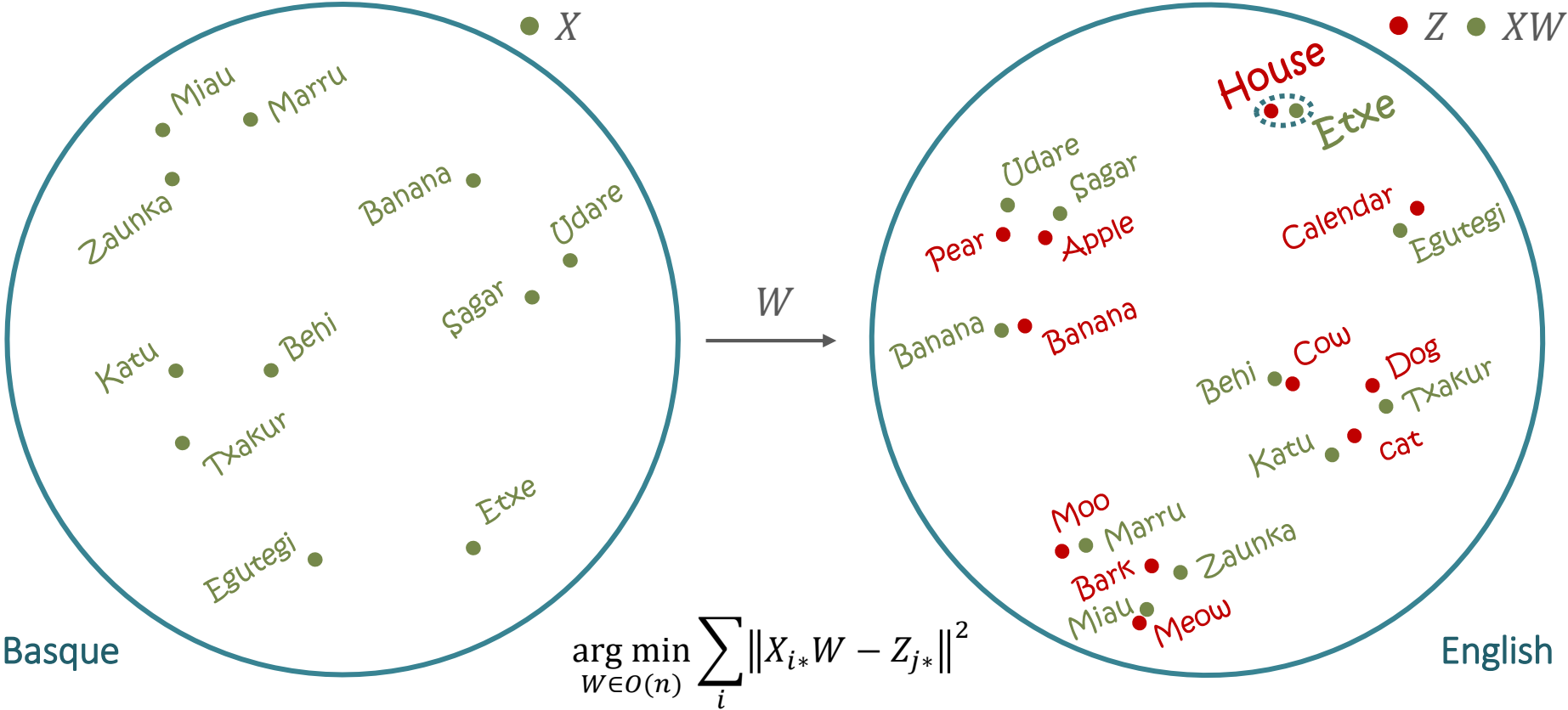
| | | | | | |
|----------|---|-------|-----------|---|----------|
| Txakur | $\begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix}$ | $[W]$ | \approx | $\begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix}$ | Dog |
| Sagar | | | | | Apple |
| \vdots | | | | | \vdots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

Bilingual embedding mappings



| | | | | | | |
|---------|---|-----|---|---|------------------|----------|
| Txakur |] | [W] | ≈ |] | Dog | |
| Sagar | | | | | Z _{2,*} | Apple |
| ⋮ | | | | | ⋮ | ⋮ |
| Egutegi | | | | | Z _{n,*} | Calendar |

Bilingual embedding mappings



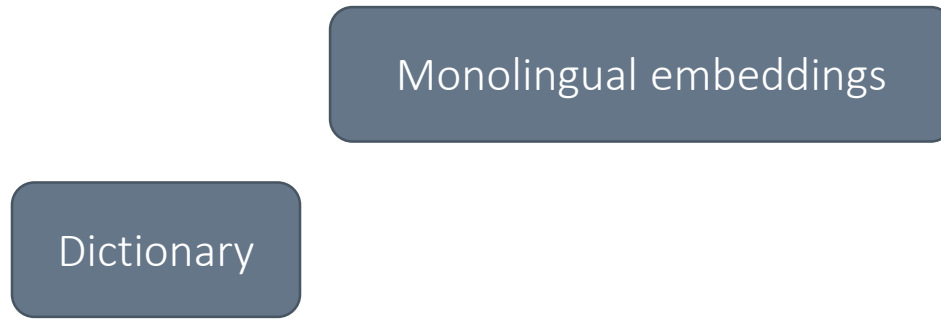
$$\begin{array}{l}
 \text{Txakur} \\
 \text{Sagar} \\
 \vdots \\
 \text{Egutegi}
 \end{array}
 \begin{bmatrix}
 X_{1,*} \\
 X_{2,*} \\
 \vdots \\
 X_{n,*}
 \end{bmatrix}
 [W] \approx
 \begin{bmatrix}
 Z_{1,*} \\
 Z_{2,*} \\
 \vdots \\
 Z_{n,*}
 \end{bmatrix}
 \begin{array}{l}
 \text{Dog} \\
 \text{Apple}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

Bilingual embedding mappings

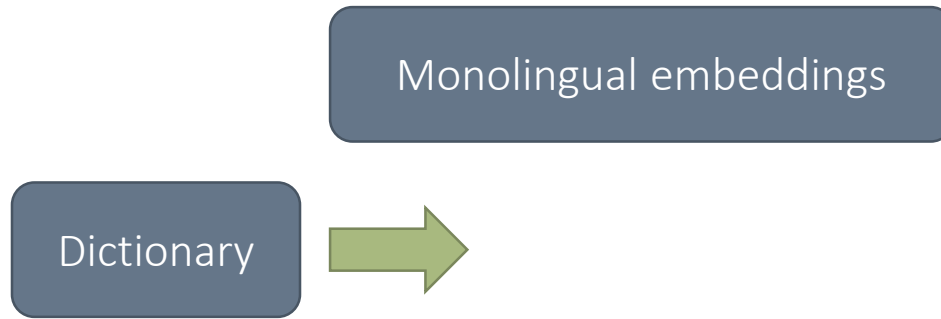
Bilingual embedding mappings

Monolingual embeddings

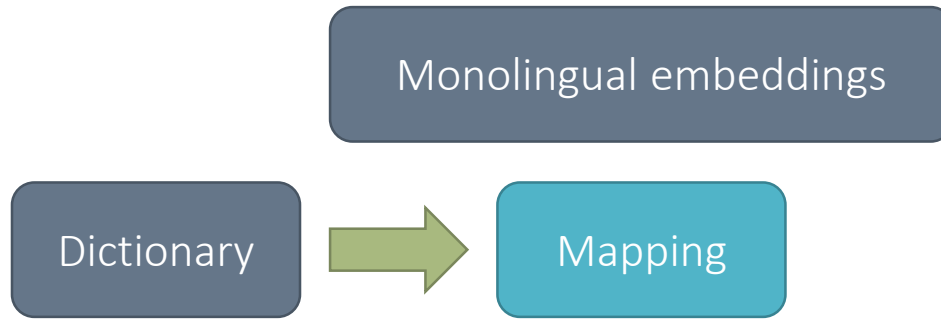
Bilingual embedding mappings



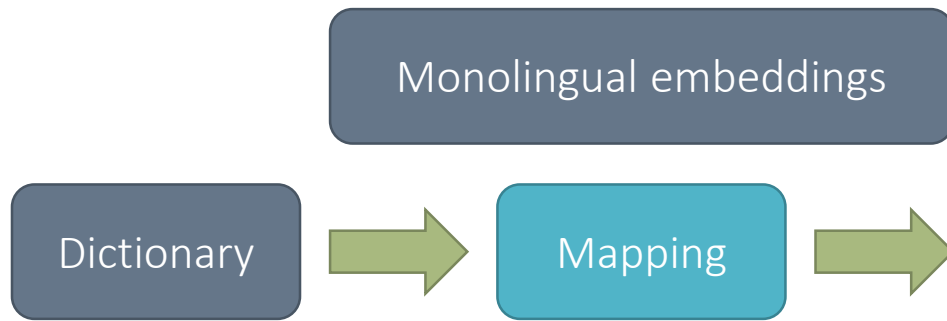
Bilingual embedding mappings



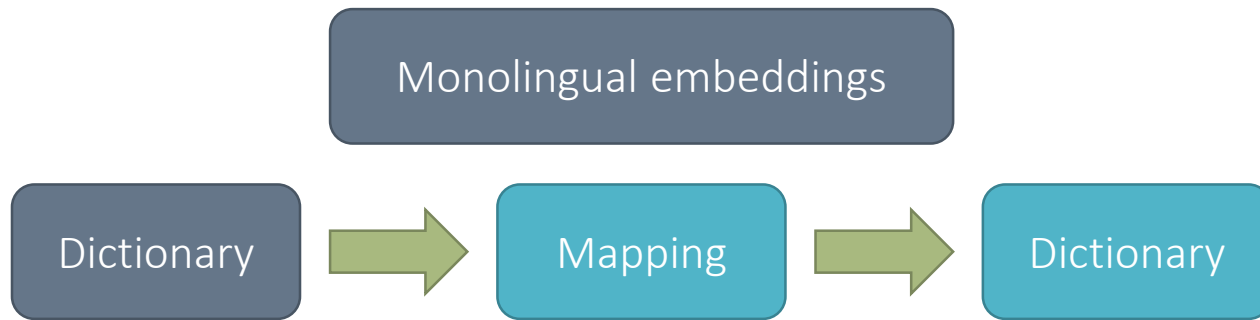
Bilingual embedding mappings



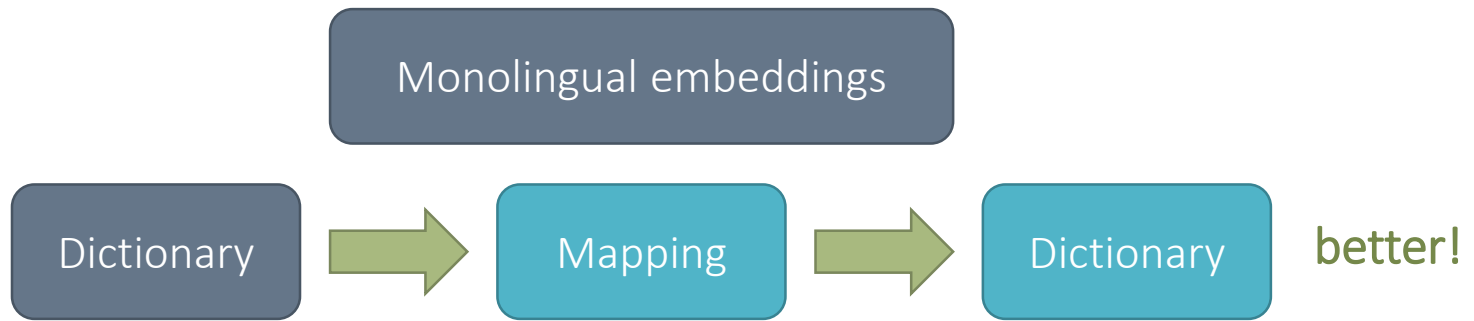
Bilingual embedding mappings



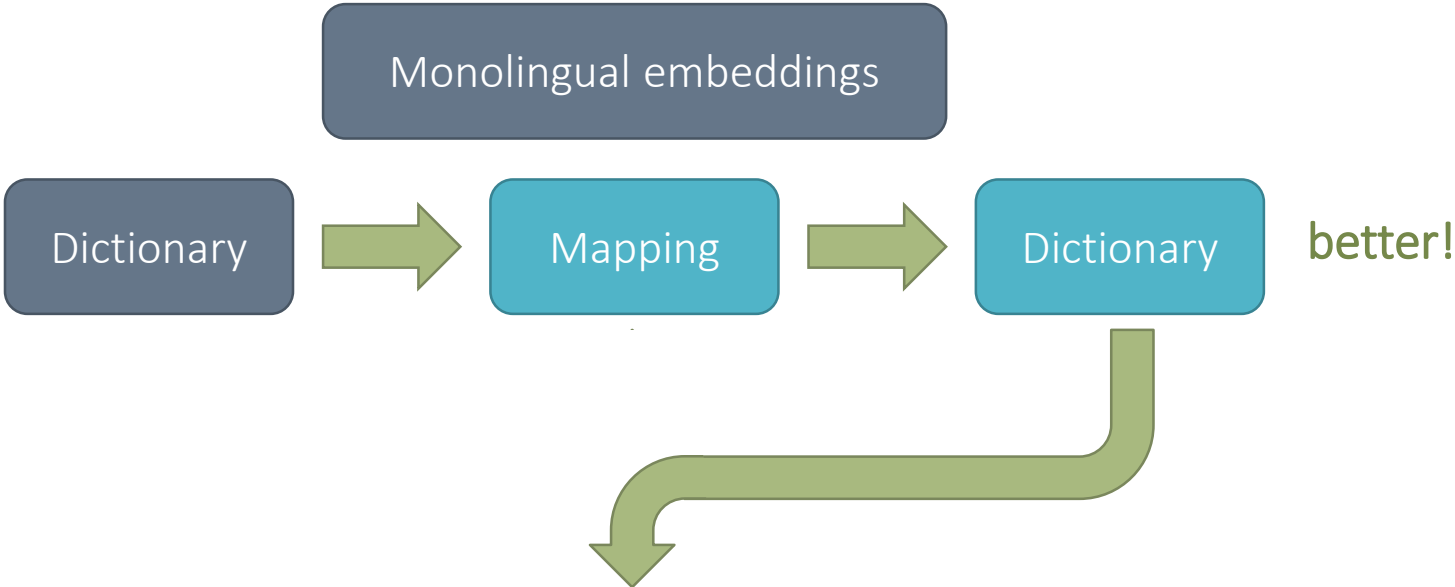
Bilingual embedding mappings



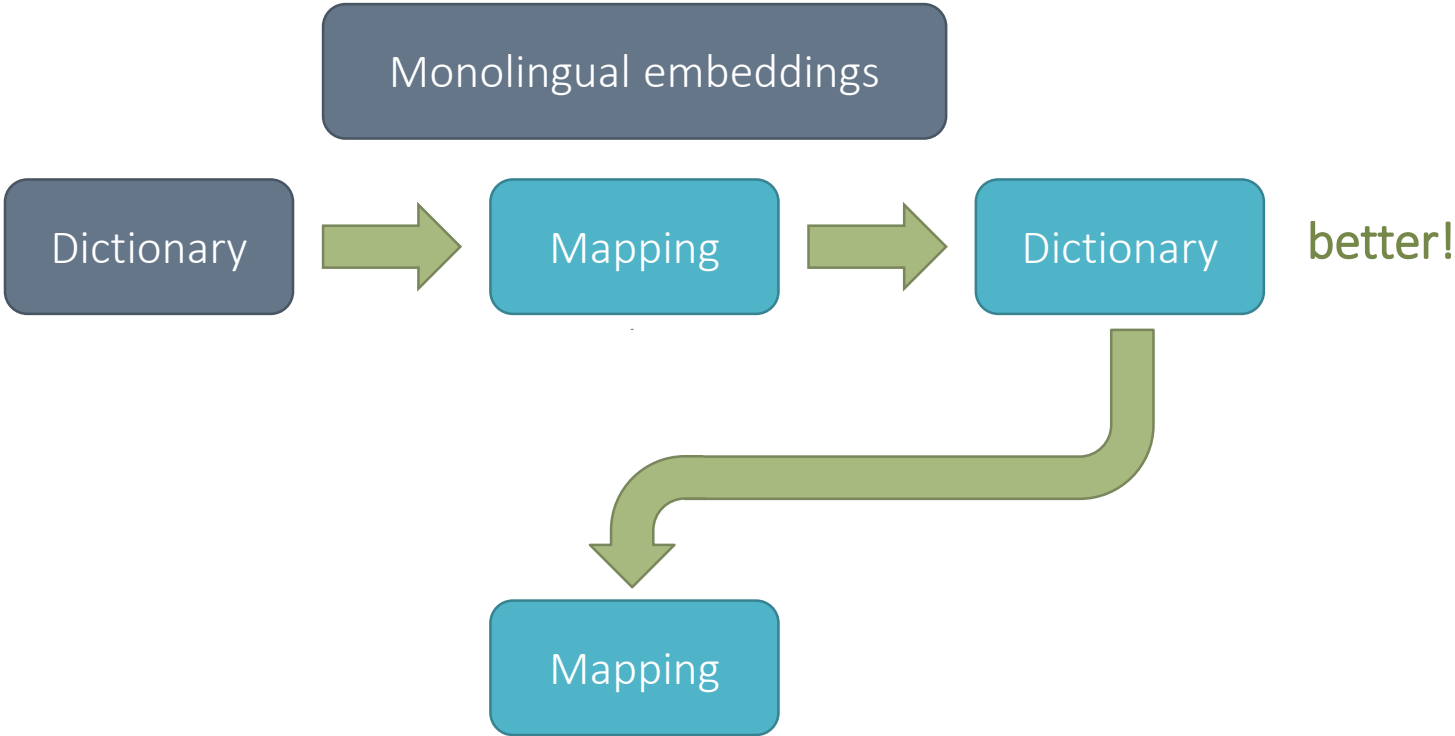
Bilingual embedding mappings



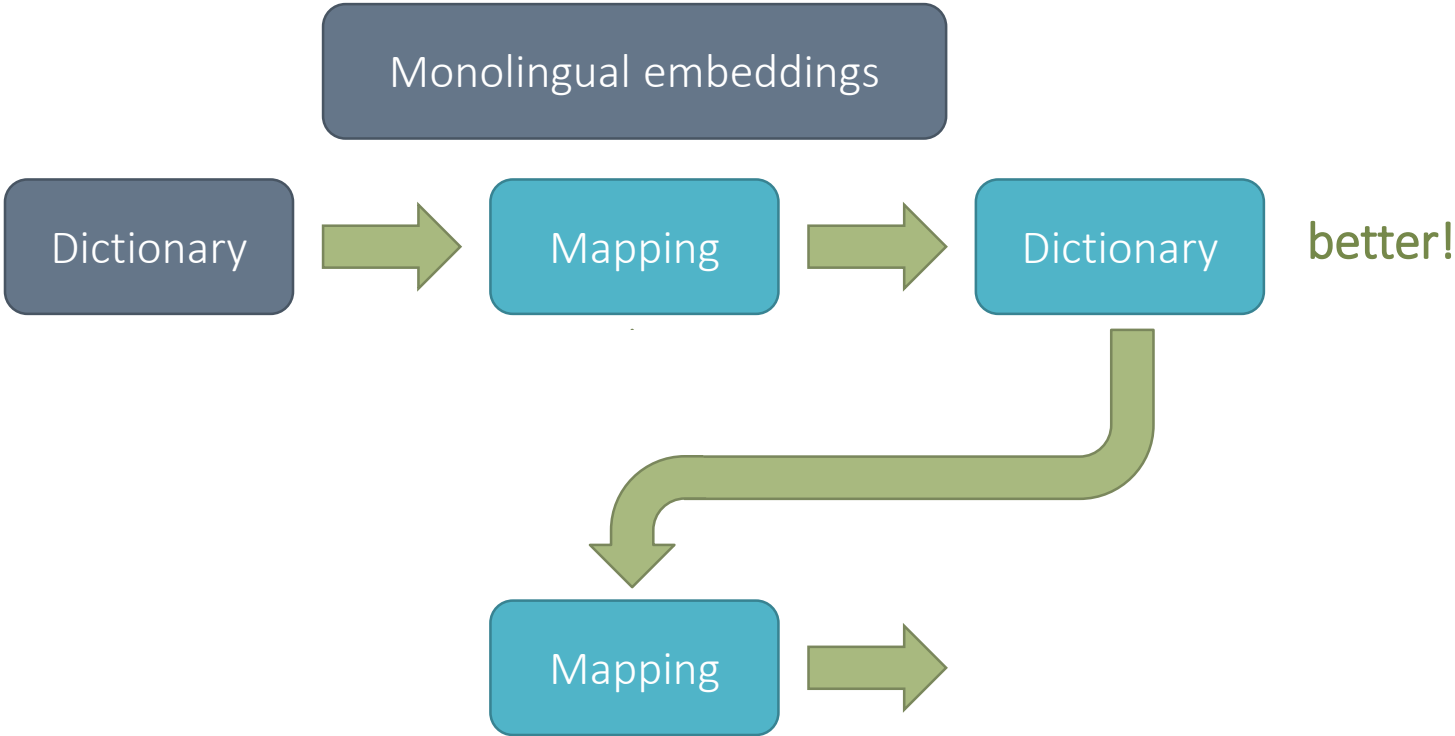
Bilingual embedding mappings



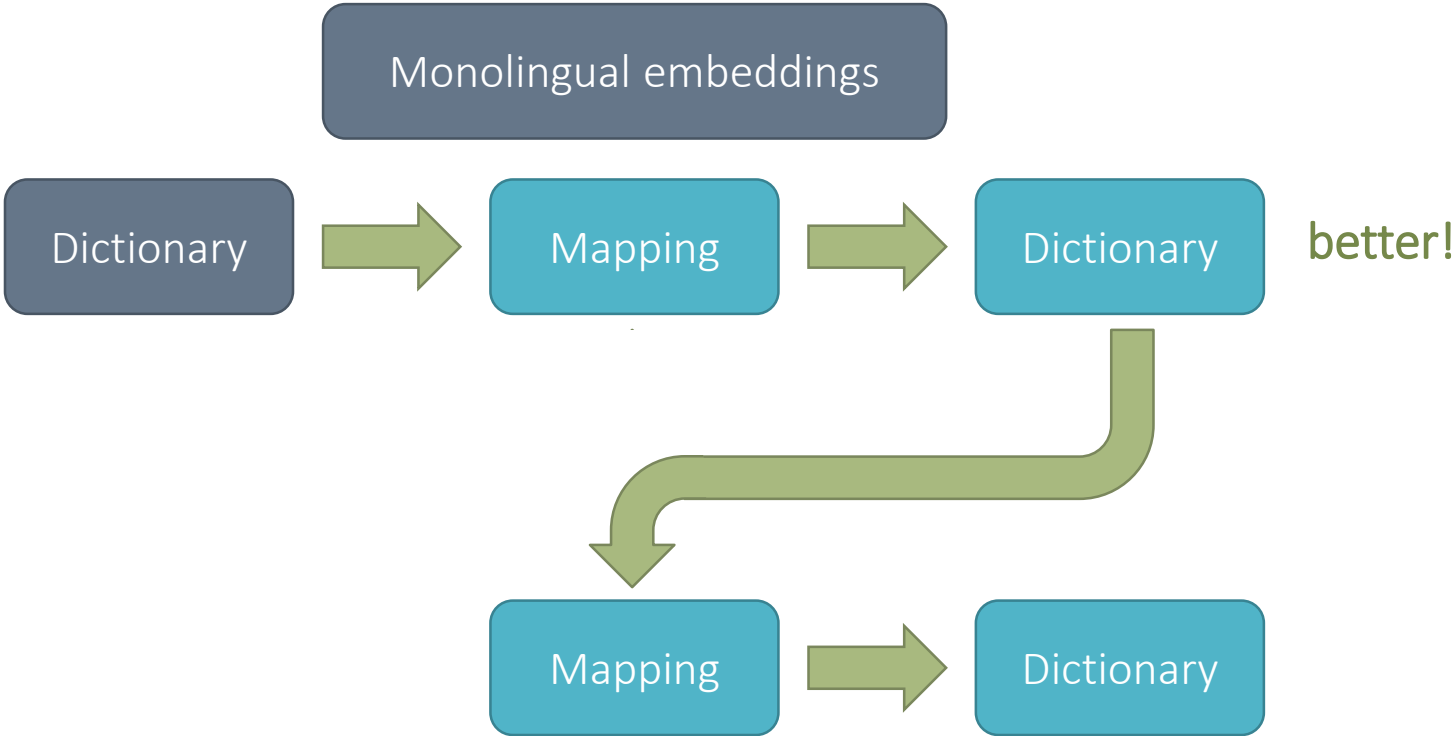
Bilingual embedding mappings



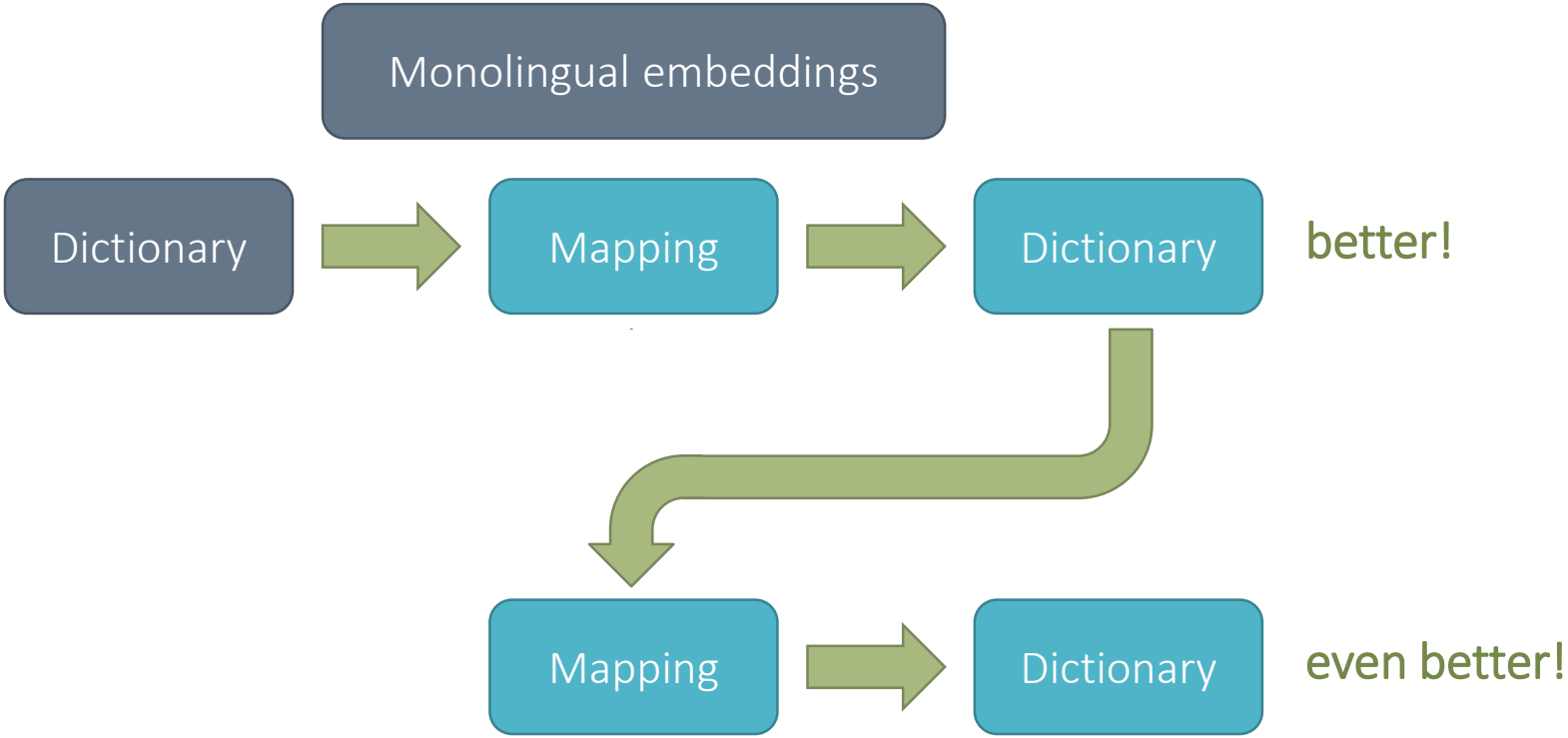
Bilingual embedding mappings



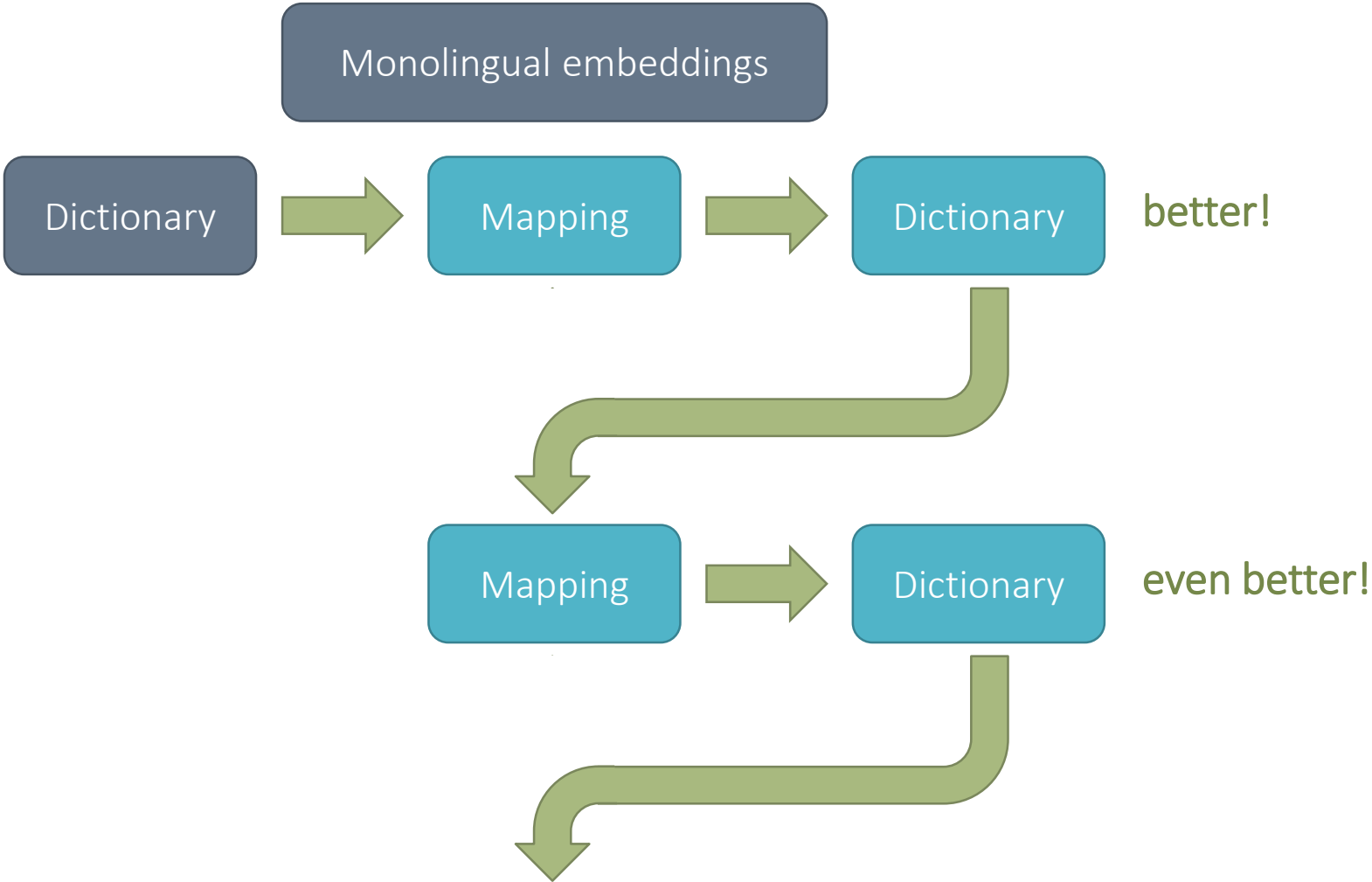
Bilingual embedding mappings



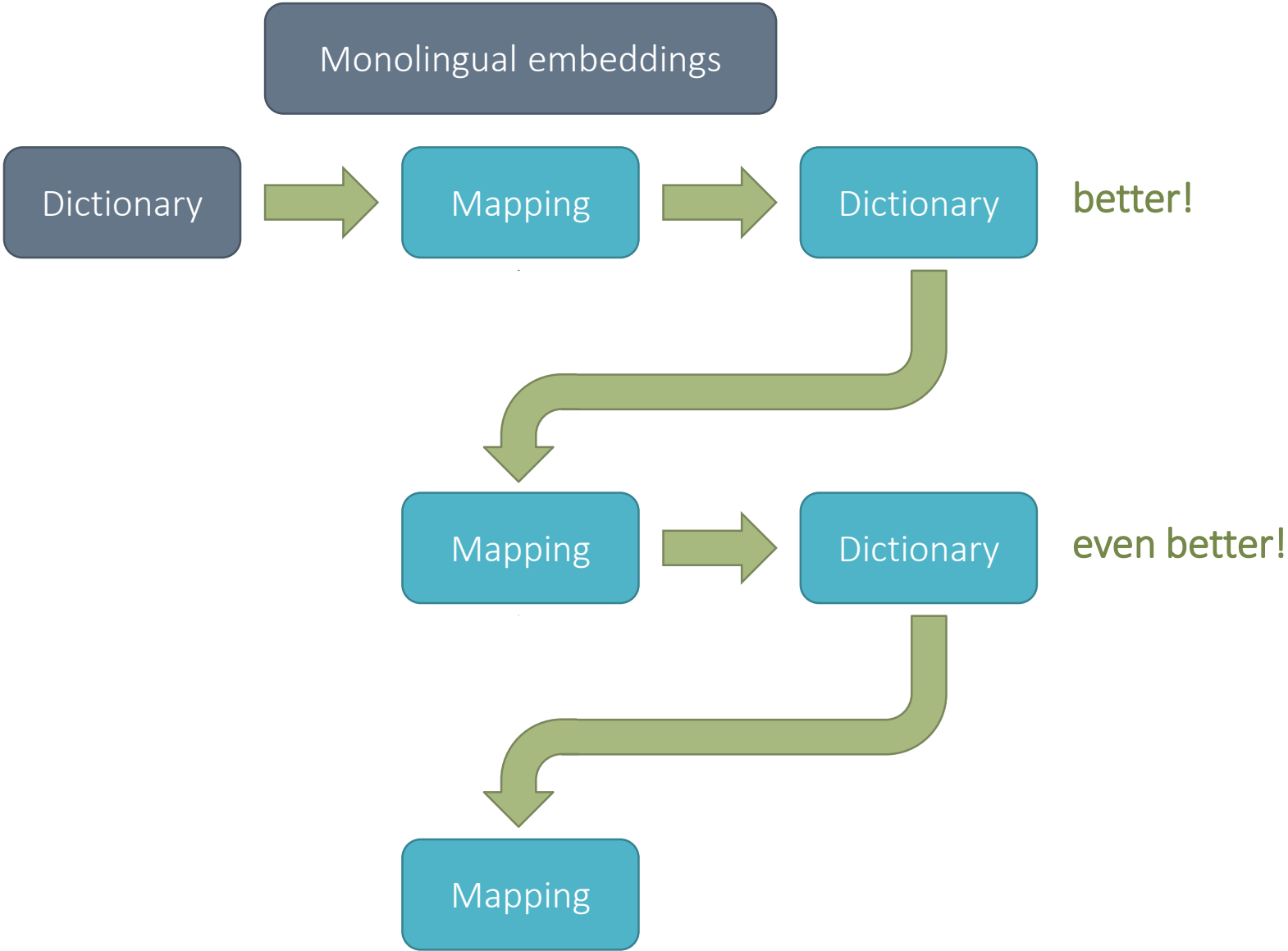
Bilingual embedding mappings



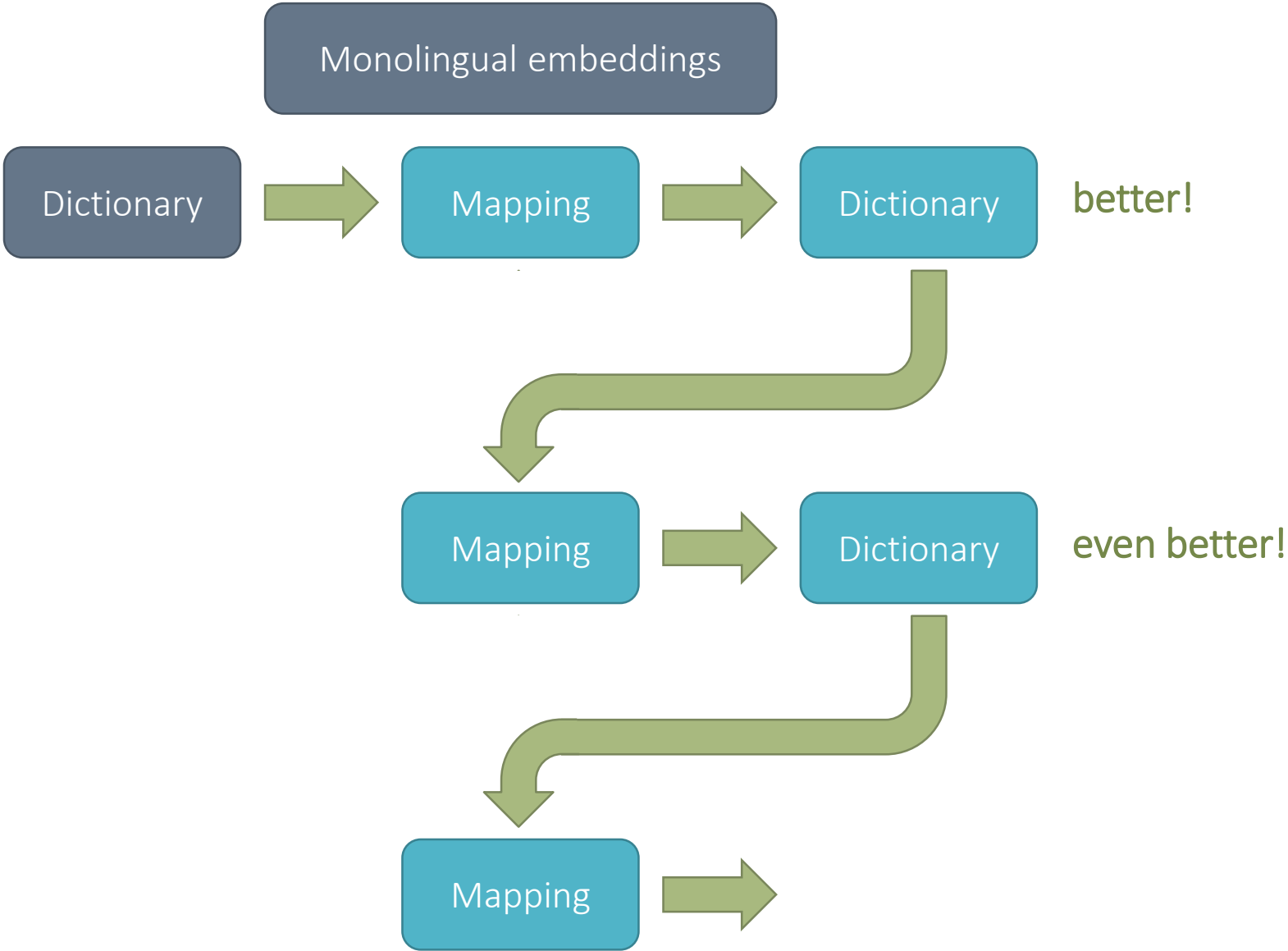
Bilingual embedding mappings



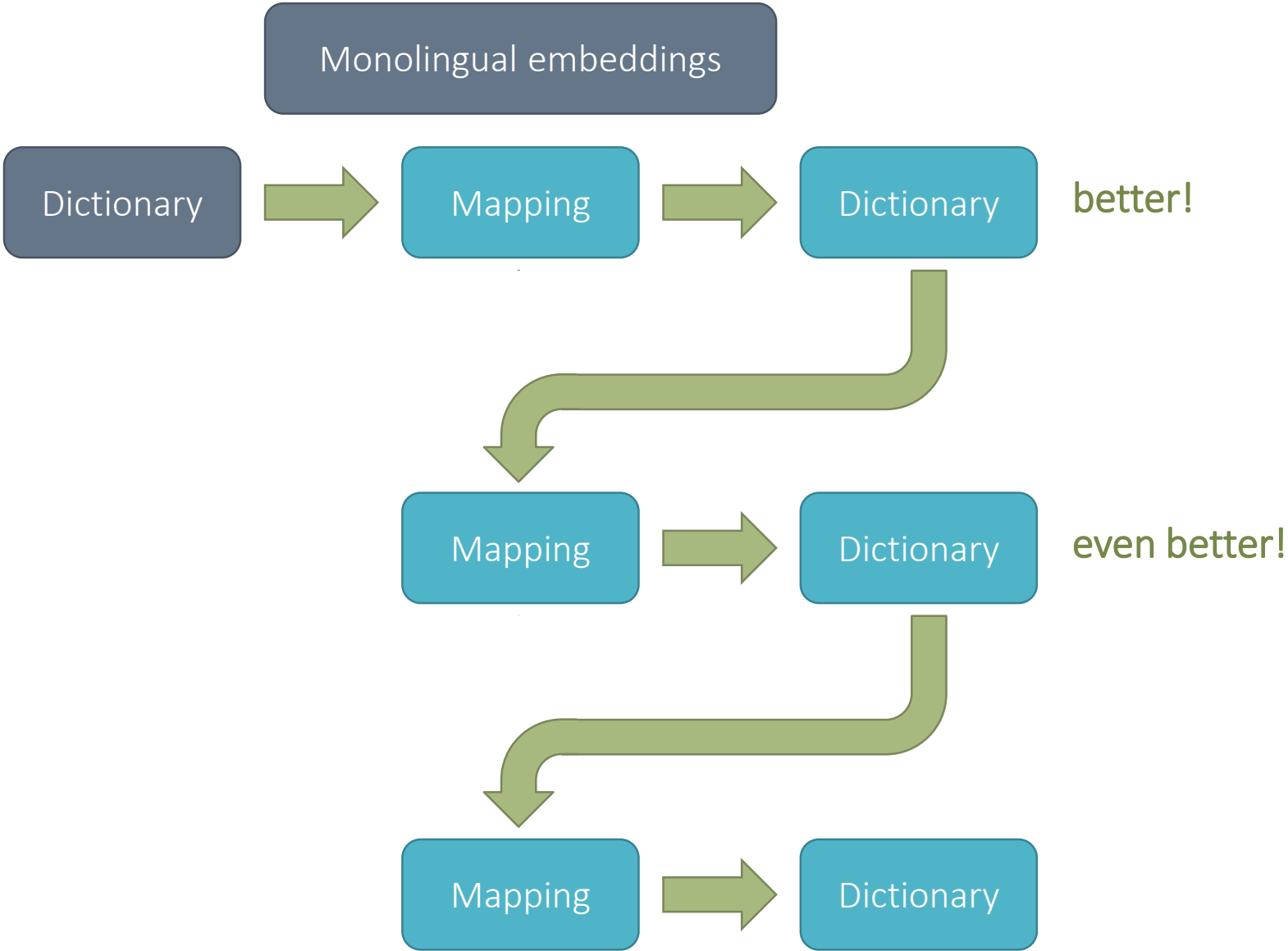
Bilingual embedding mappings



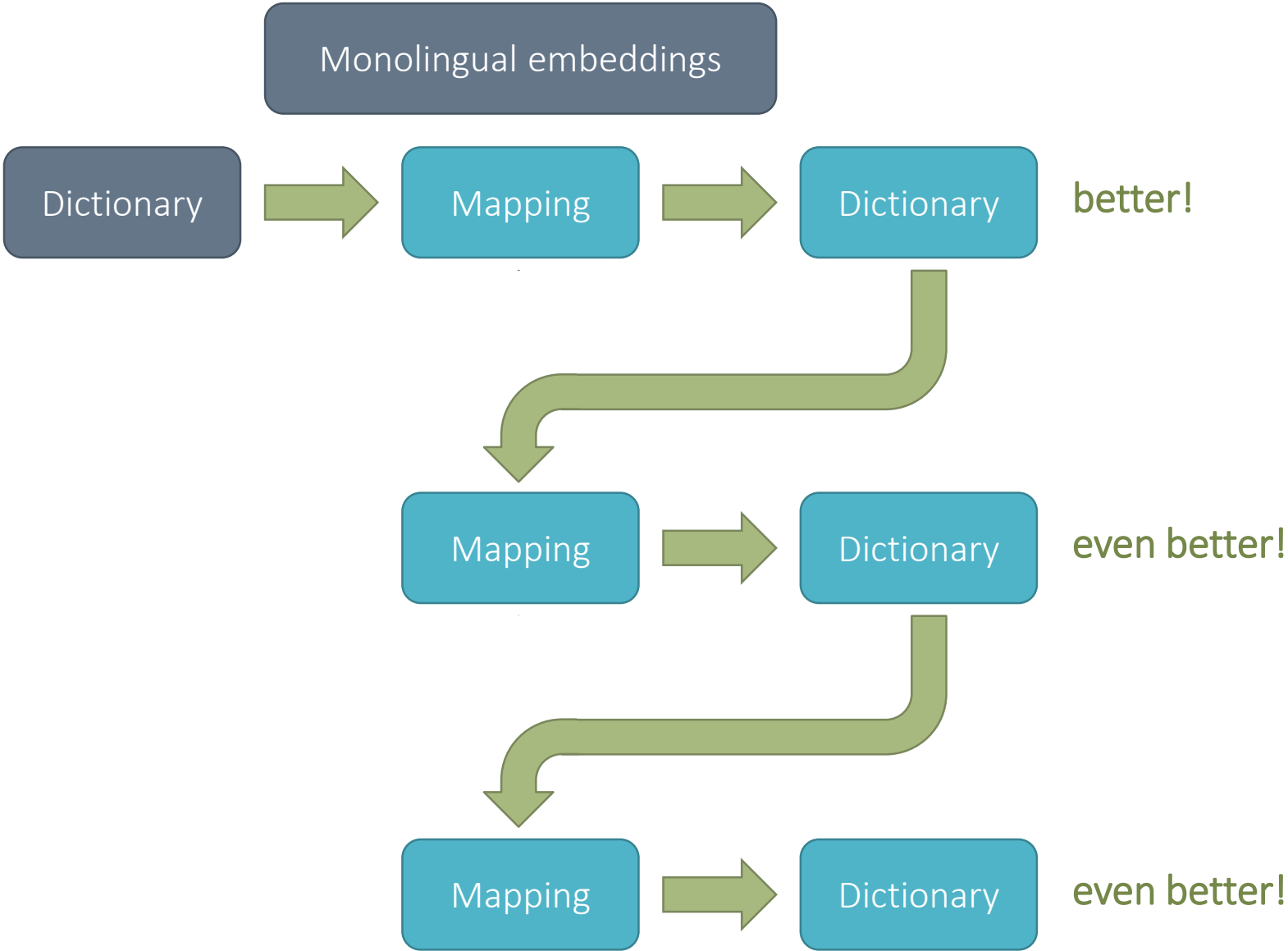
Bilingual embedding mappings



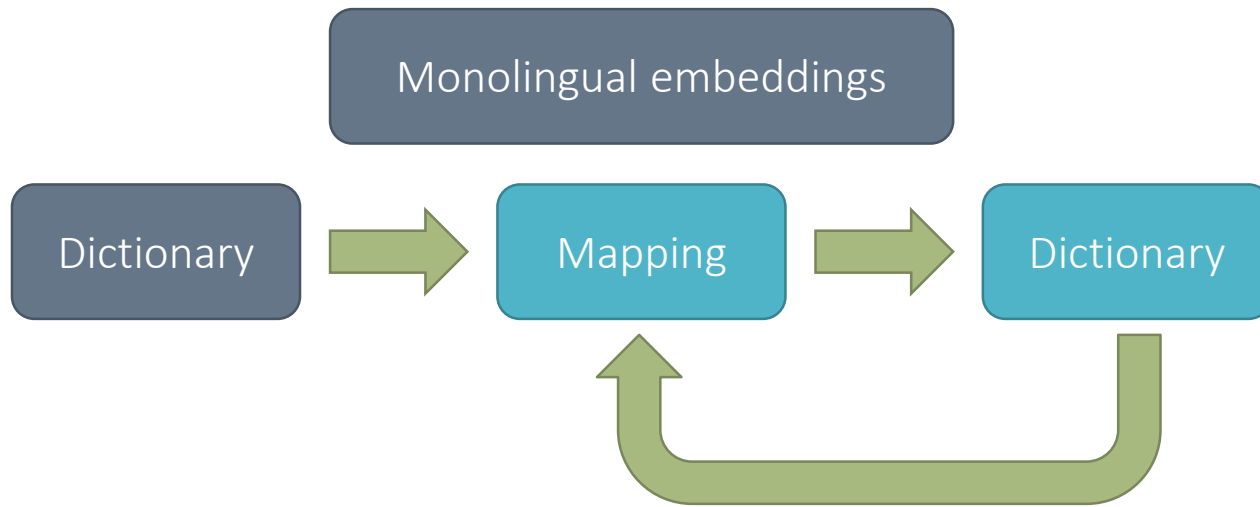
Bilingual embedding mappings



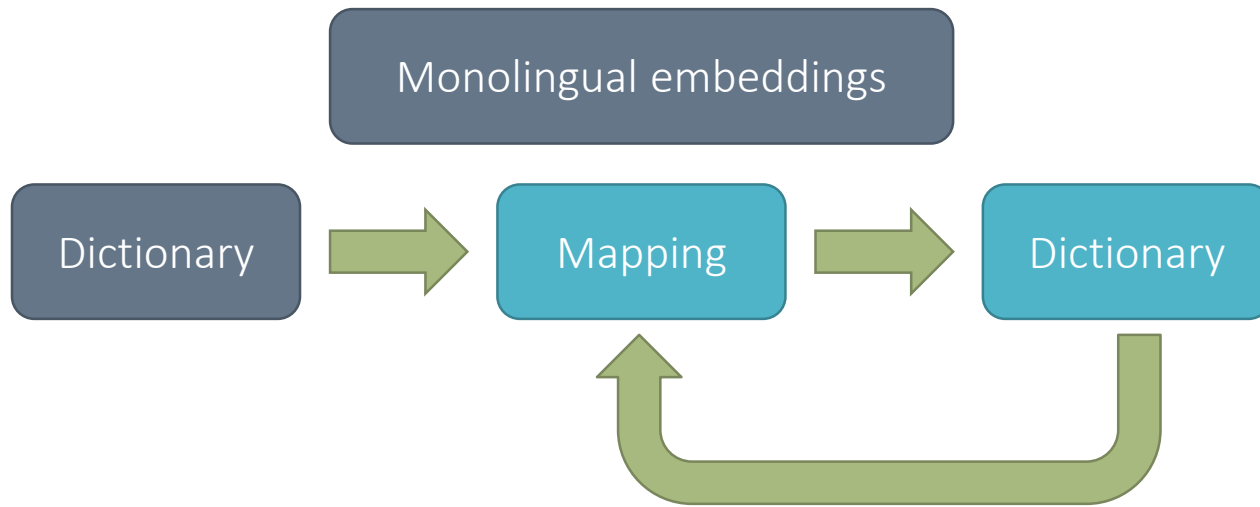
Bilingual embedding mappings



Bilingual embedding mappings

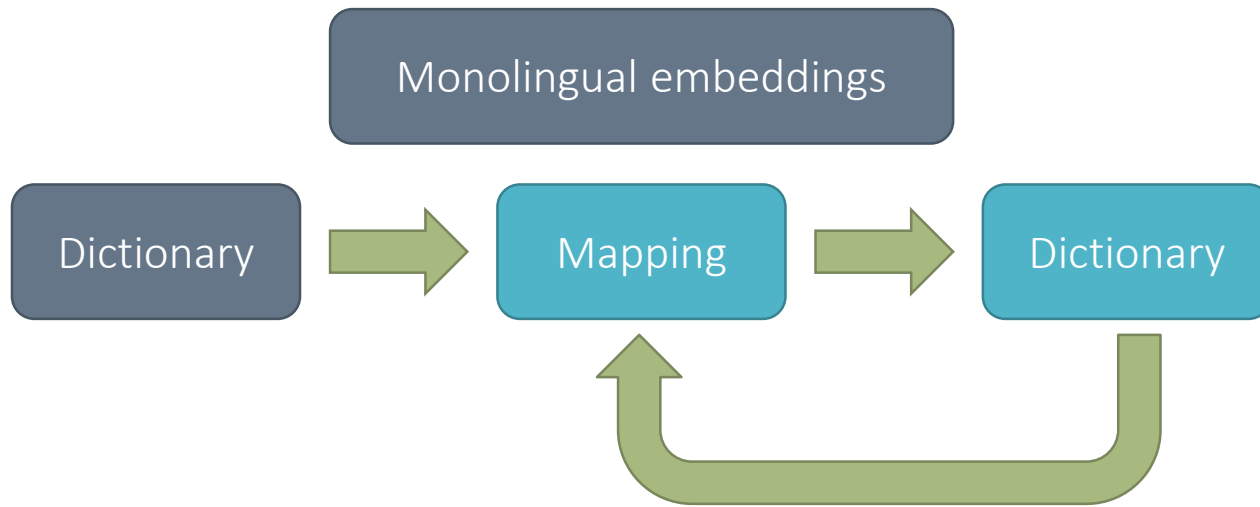


Bilingual embedding mappings



proposed self-learning method

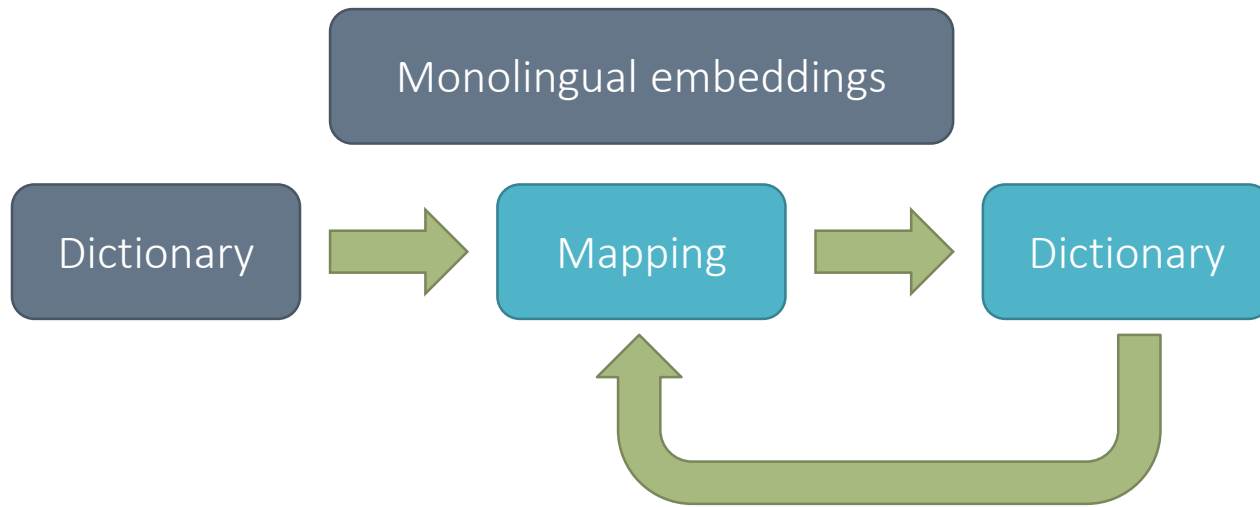
Bilingual embedding mappings



proposed self-learning method

formalization and implementation details in the paper
based on the mapping method of Artetxe et al. (2016)

Bilingual embedding mappings

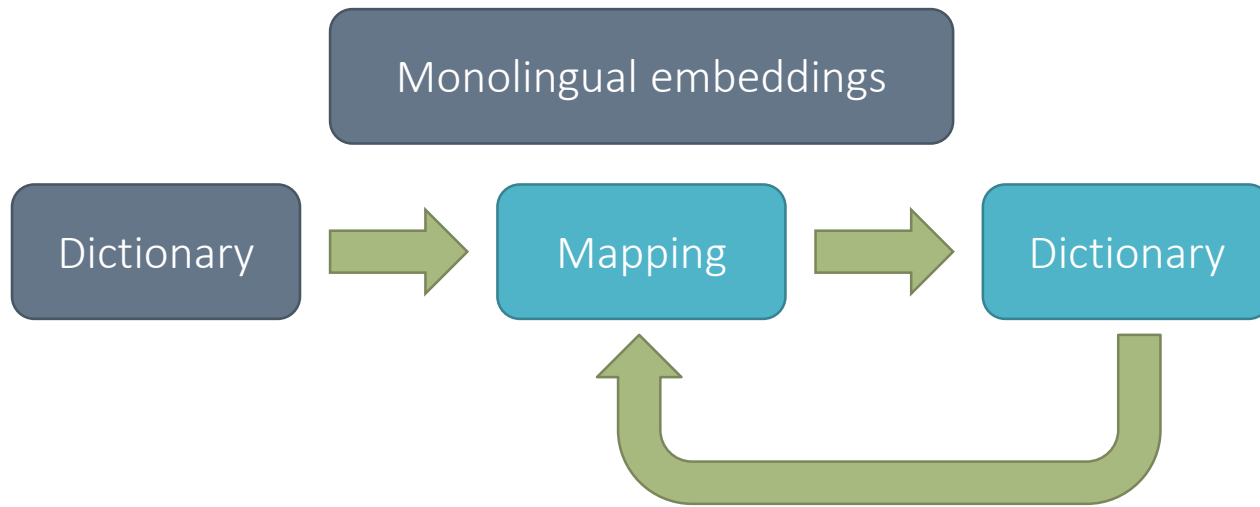


proposed self-learning method

formalization and implementation details in the paper
based on the mapping method of Artetxe et al. (2016)

Too good to be true?

Bilingual embedding mappings



proposed self-learning method

formalization and implementation details in the paper
based on the mapping method of Artetxe et al. (2016)

Too good to be true?

Nope, it works!

Experiments

Experiments

- Dataset by Dinu et al. (2015)

Experiments

- Dataset by Dinu et al. (2015)

English-Italian

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish

English-Italian

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish

| | | |
|-----------------|----------------|-----------------|
| English-Italian | English-German | English-Finnish |
|-----------------|----------------|-----------------|

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
⇒ *Monolingual embeddings (CBOW + negative sampling)*

| | | |
|-----------------|----------------|-----------------|
| English-Italian | English-German | English-Finnish |
|-----------------|----------------|-----------------|

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs*

| English-Italian | | English-German | | English-Finnish | |
|-----------------|--|----------------|--|-----------------|--|
| 5,000 | | 5,000 | | 5,000 | |

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs*

| English-Italian | | English-German | | English-Finnish | |
|-----------------|----|----------------|----|-----------------|----|
| 5,000 | 25 | 5,000 | 25 | 5,000 | 25 |

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

| English-Italian | | | English-German | | | English-Finnish | | |
|-----------------|----|------|----------------|----|------|-----------------|----|------|
| 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*
 - ⇒ *Test dictionary: 1,500 word pairs*

| English-Italian | | | English-German | | | English-Finnish | | |
|-----------------|----|------|----------------|----|------|-----------------|----|------|
| 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*
 - ⇒ *Test dictionary: 1,500 word pairs*

| English-Italian | | | English-German | | | English-Finnish | | |
|-----------------|----|------|----------------|----|------|-----------------|----|------|
| 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| English-Italian | | | English-German | | | English-Finnish | | |
|-----------------|----|------|----------------|----|------|-----------------|----|------|
| 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |

Mikolov et al. (2013a)

Xing et al. (2015)

Zhang et al. (2016)

Artetxe et al. (2016)

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| English-Italian | | | English-German | | | English-Finnish | | |
|-----------------|----|------|----------------|----|------|-----------------|----|------|
| 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |

| |
|------------------------|
| Mikolov et al. (2013a) |
| Xing et al. (2015) |
| Zhang et al. (2016) |
| Artetxe et al. (2016) |
| Our method |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| | English-Italian | | | English-German | | | English-Finnish | | |
|------------------------|-----------------|---------------|---------------|----------------|---------------|---------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% | 35.00% | 0.00% | 0.07% | 25.91% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% | 41.27% | 0.07% | 0.53% | 28.23% | 0.07% | 0.56% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% | 40.80% | 0.13% | 0.87% | 28.16% | 0.14% | 0.42% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% | 41.87% | 0.13% | 0.73% | 30.62% | 0.21% | 0.77% |
| Our method | 39.67% | 37.27% | 39.40% | 40.87% | 39.60% | 40.27% | 28.72% | 28.16% | 26.47% |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| | English-Italian | | | English-German | | | English-Finnish | | |
|------------------------|-----------------|---------------|---------------|----------------|---------------|---------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% | 35.00% | 0.00% | 0.07% | 25.91% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% | 41.27% | 0.07% | 0.53% | 28.23% | 0.07% | 0.56% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% | 40.80% | 0.13% | 0.87% | 28.16% | 0.14% | 0.42% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% | 41.87% | 0.13% | 0.73% | 30.62% | 0.21% | 0.77% |
| Our method | 39.67% | 37.27% | 39.40% | 40.87% | 39.60% | 40.27% | 28.72% | 28.16% | 26.47% |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| | English-Italian | | | English-German | | | English-Finnish | | |
|------------------------|-----------------|---------------|---------------|----------------|---------------|---------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% | 35.00% | 0.00% | 0.07% | 25.91% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% | 41.27% | 0.07% | 0.53% | 28.23% | 0.07% | 0.56% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% | 40.80% | 0.13% | 0.87% | 28.16% | 0.14% | 0.42% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% | 41.87% | 0.13% | 0.73% | 30.62% | 0.21% | 0.77% |
| Our method | 39.67% | 37.27% | 39.40% | 40.87% | 39.60% | 40.27% | 28.72% | 28.16% | 26.47% |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| | English-Italian | | | English-German | | | English-Finnish | | |
|------------------------|-----------------|---------------|---------------|----------------|---------------|---------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% | 35.00% | 0.00% | 0.07% | 25.91% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% | 41.27% | 0.07% | 0.53% | 28.23% | 0.07% | 0.56% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% | 40.80% | 0.13% | 0.87% | 28.16% | 0.14% | 0.42% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% | 41.87% | 0.13% | 0.73% | 30.62% | 0.21% | 0.77% |
| Our method | 39.67% | 37.27% | 39.40% | 40.87% | 39.60% | 40.27% | 28.72% | 28.16% | 26.47% |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

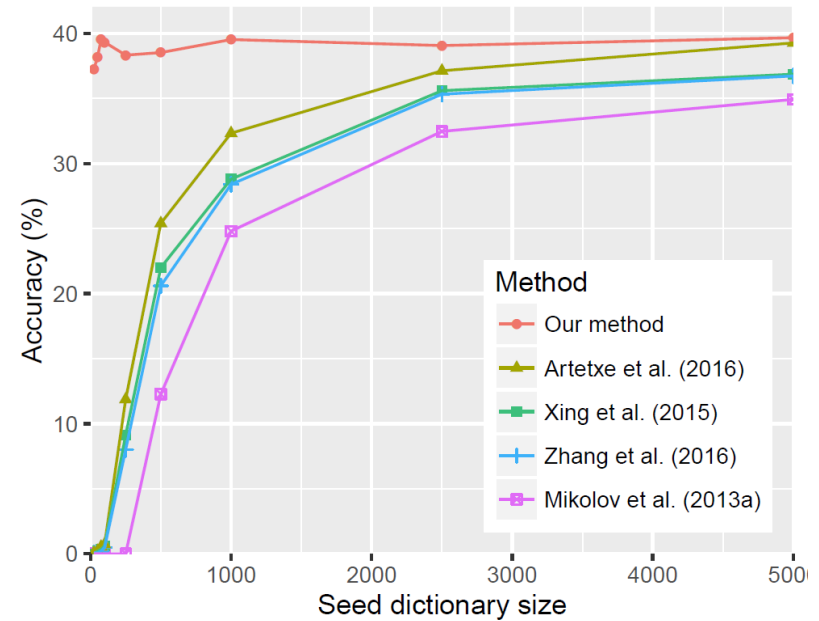
| | English-Italian | | | English-German | | | English-Finnish | | |
|------------------------|-----------------|---------------|---------------|----------------|---------------|---------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. | 5,000 | 25 | num. | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% | 35.00% | 0.00% | 0.07% | 25.91% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% | 41.27% | 0.07% | 0.53% | 28.23% | 0.07% | 0.56% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% | 40.80% | 0.13% | 0.87% | 28.16% | 0.14% | 0.42% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% | 41.87% | 0.13% | 0.73% | 30.62% | 0.21% | 0.77% |
| Our method | 39.67% | 37.27% | 39.40% | 40.87% | 39.60% | 40.27% | 28.72% | 28.16% | 26.47% |

word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals
 - ⇒ Test dictionary: 1,500 word pairs

| | English-Italian | | |
|------------------------|-----------------|---------------|---------------|
| | 5,000 | 25 | num. |
| Mikolov et al. (2013a) | 34.93% | 0.00% | 0.00% |
| Xing et al. (2015) | 36.87% | 0.00% | 0.13% |
| Zhang et al. (2016) | 36.73% | 0.07% | 0.27% |
| Artetxe et al. (2016) | 39.27% | 0.07% | 0.40% |
| Our method | 39.67% | 37.27% | 39.40% |



word translation induction

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

| | | |
|-------|-------|----|
| EN-IT | EN-DE | |
| WS | RG | WS |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

| | | | |
|----------|-------|-------|----|
| | EN-IT | EN-DE | |
| Bi. data | WS | RG | WS |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ *Monolingual embeddings (CBOW + negative sampling)*
 - ⇒ *Seed dictionary: 5,000 word pairs / 25 word pairs / numerals*

| | | | | |
|---------------------|----------|-------|-------|----|
| | | EN-IT | EN-DE | |
| | Bi. data | WS | RG | WS |
| Luong et al. (2015) | Europarl | | | |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals

| | | EN-IT | EN-DE | |
|------------------------|----------|-------|-------|----|
| | | WS | RG | WS |
| Luong et al. (2015) | Europarl | | | |
| Mikolov et al. (2013a) | 5k dict | | | |
| Xing et al. (2015) | 5k dict | | | |
| Zhang et al. (2016) | 5k dict | | | |
| Artetxe et al. (2016) | 5k dict | | | |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals

| | | EN-IT | EN-DE | |
|------------------------|----------------------------|-------|-------|----|
| | | WS | RG | WS |
| Luong et al. (2015) | Europarl | | | |
| Mikolov et al. (2013a) | 5k dict | | | |
| Xing et al. (2015) | 5k dict | | | |
| Zhang et al. (2016) | 5k dict | | | |
| Artetxe et al. (2016) | 5k dict | | | |
| Our method | 5k dict 25 dict num. | | | |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals

| | Bi. data | EN-IT | EN-DE | |
|------------------------|----------|--------------|--------------|--------------|
| | | WS | RG | WS |
| Luong et al. (2015) | Europarl | 33.1% | 33.5% | 35.6% |
| Mikolov et al. (2013a) | 5k dict | 62.7% | 64.3% | 52.8% |
| Xing et al. (2015) | 5k dict | 61.4% | 70.0% | 59.5% |
| Zhang et al. (2016) | 5k dict | 61.6% | 70.4% | 59.6% |
| Artetxe et al. (2016) | 5k dict | 61.7% | 71.6% | 59.7% |
| Our method | 5k dict | 62.4% | 74.2% | 61.6% |
| | 25 dict | 62.6% | 74.9% | 61.2% |
| | num. | 62.8% | 73.9% | 60.4% |

crosslingual word similarity

Experiments

- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals

| | Bi. data | EN-IT | EN-DE | |
|------------------------|----------|--------------|--------------|--------------|
| | | WS | RG | WS |
| Luong et al. (2015) | Europarl | 33.1% | 33.5% | 35.6% |
| Mikolov et al. (2013a) | 5k dict | 62.7% | 64.3% | 52.8% |
| Xing et al. (2015) | 5k dict | 61.4% | 70.0% | 59.5% |
| Zhang et al. (2016) | 5k dict | 61.6% | 70.4% | 59.6% |
| Artetxe et al. (2016) | 5k dict | 61.7% | 71.6% | 59.7% |
| Our method | 5k dict | 62.4% | 74.2% | 61.6% |
| | 25 dict | 62.6% | 74.9% | 61.2% |
| | num. | 62.8% | 73.9% | 60.4% |

crosslingual word similarity

Experiments

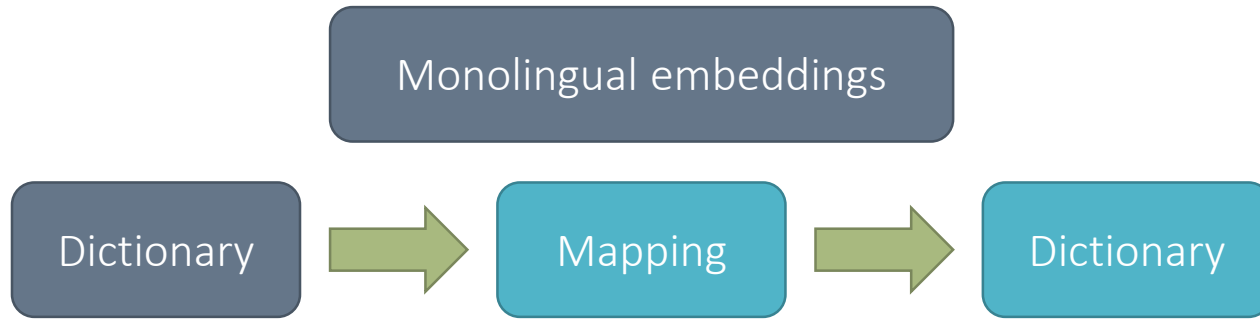
- Dataset by Dinu et al. (2015) extended to German and Finnish
 - ⇒ Monolingual embeddings (CBOW + negative sampling)
 - ⇒ Seed dictionary: 5,000 word pairs / 25 word pairs / numerals

| | Bi. data | EN-IT | EN-DE | |
|------------------------|----------|--------------|--------------|--------------|
| | | WS | RG | WS |
| Luong et al. (2015) | Europarl | 33.1% | 33.5% | 35.6% |
| Mikolov et al. (2013a) | 5k dict | 62.7% | 64.3% | 52.8% |
| Xing et al. (2015) | 5k dict | 61.4% | 70.0% | 59.5% |
| Zhang et al. (2016) | 5k dict | 61.6% | 70.4% | 59.6% |
| Artetxe et al. (2016) | 5k dict | 61.7% | 71.6% | 59.7% |
| Our method | 5k dict | 62.4% | 74.2% | 61.6% |
| | 25 dict | 62.6% | 74.9% | 61.2% |
| | num. | 62.8% | 73.9% | 60.4% |

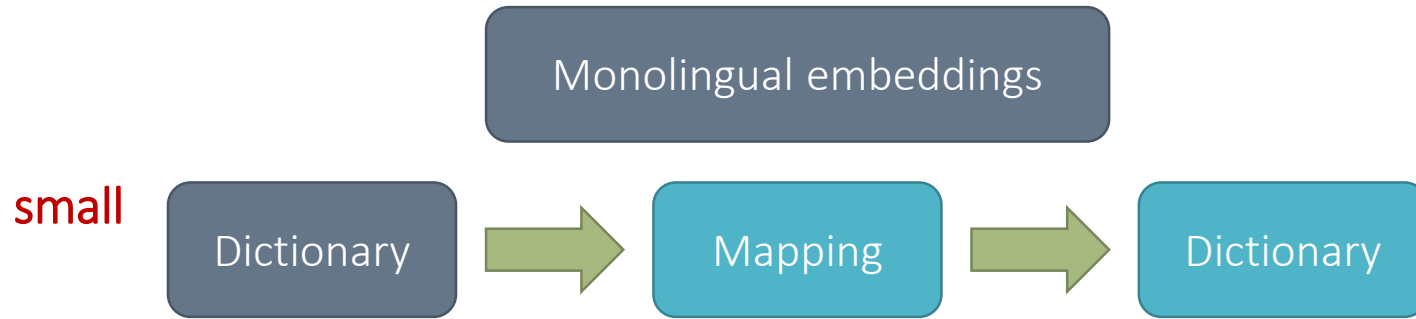
crosslingual word similarity

Why does it work?

Why does it work?



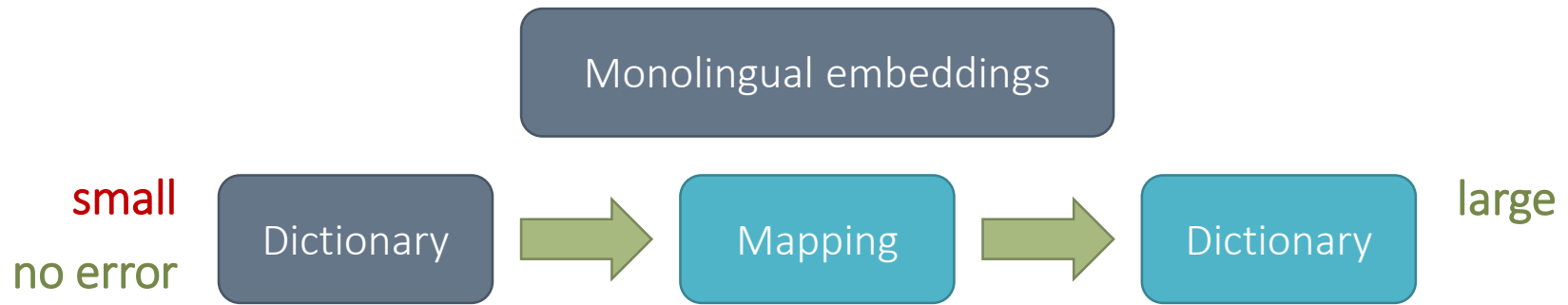
Why does it work?



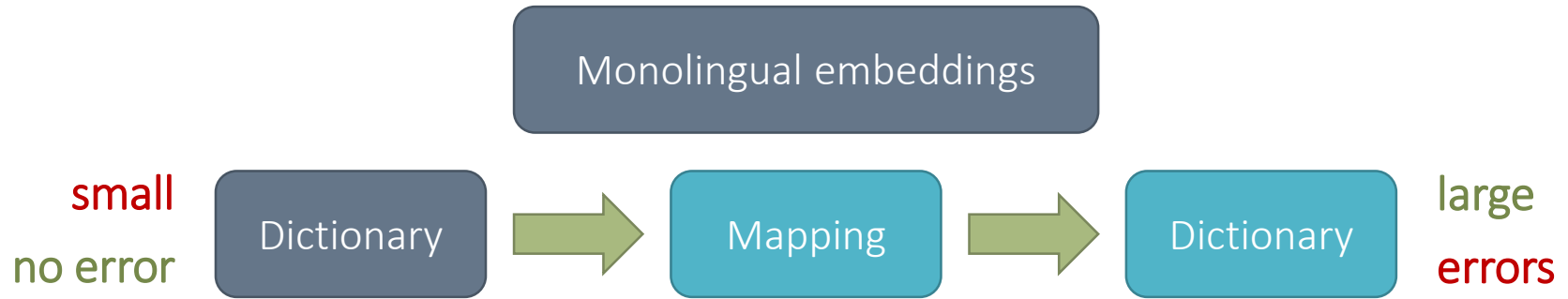
Why does it work?



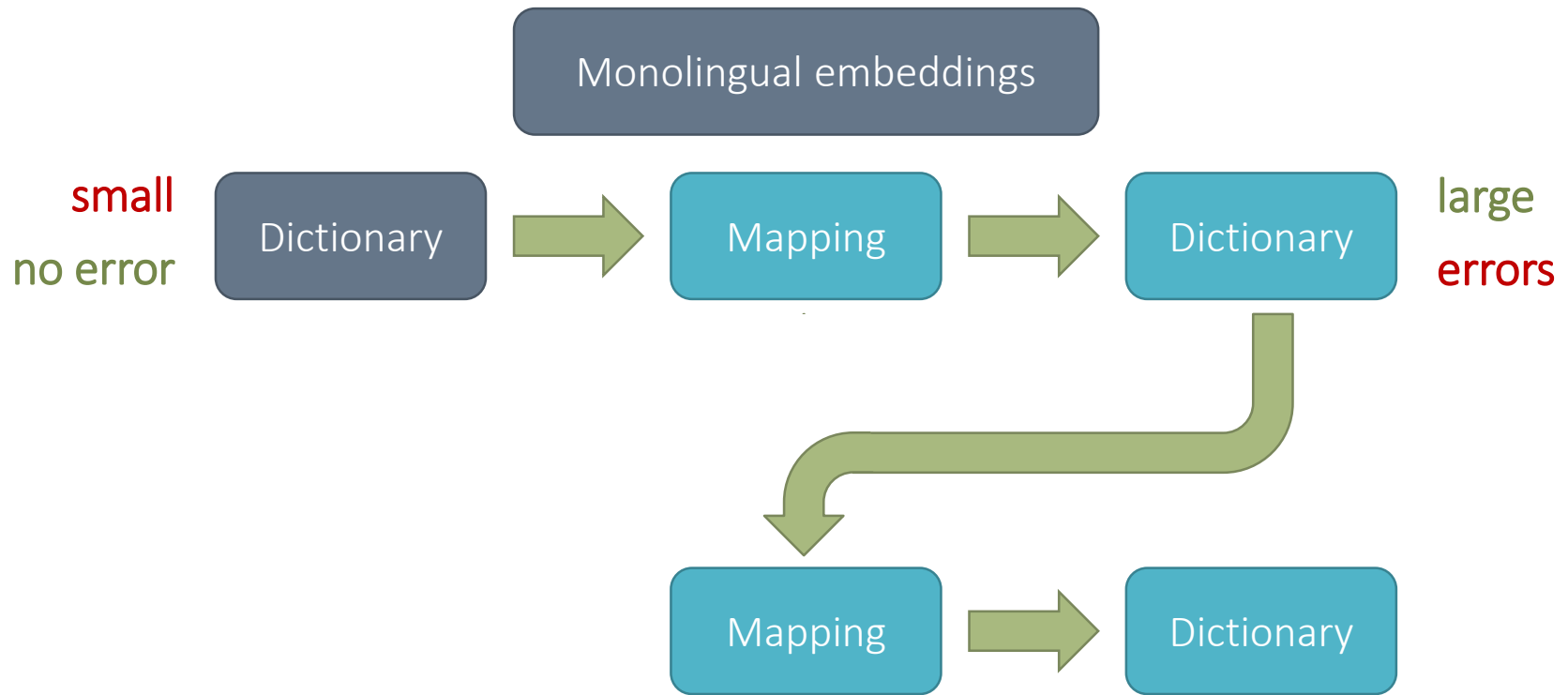
Why does it work?



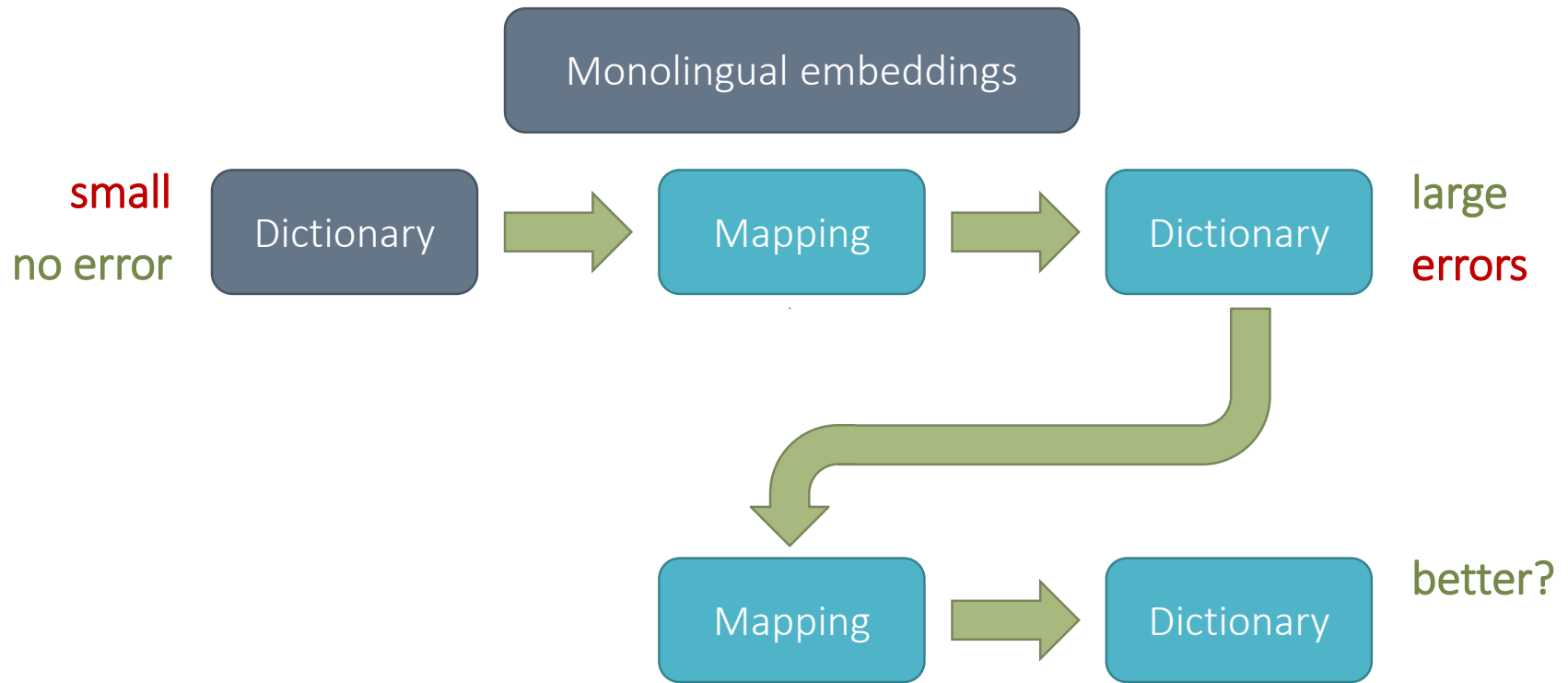
Why does it work?



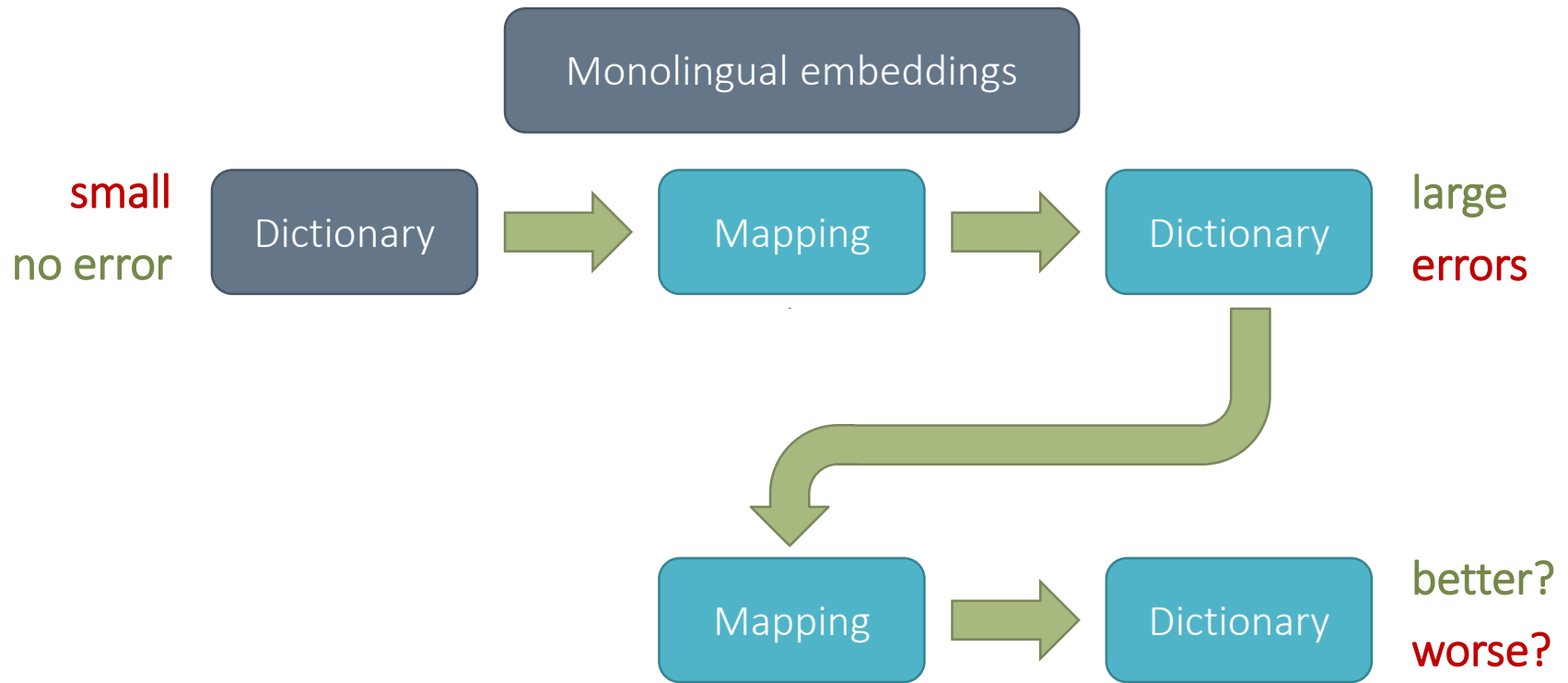
Why does it work?



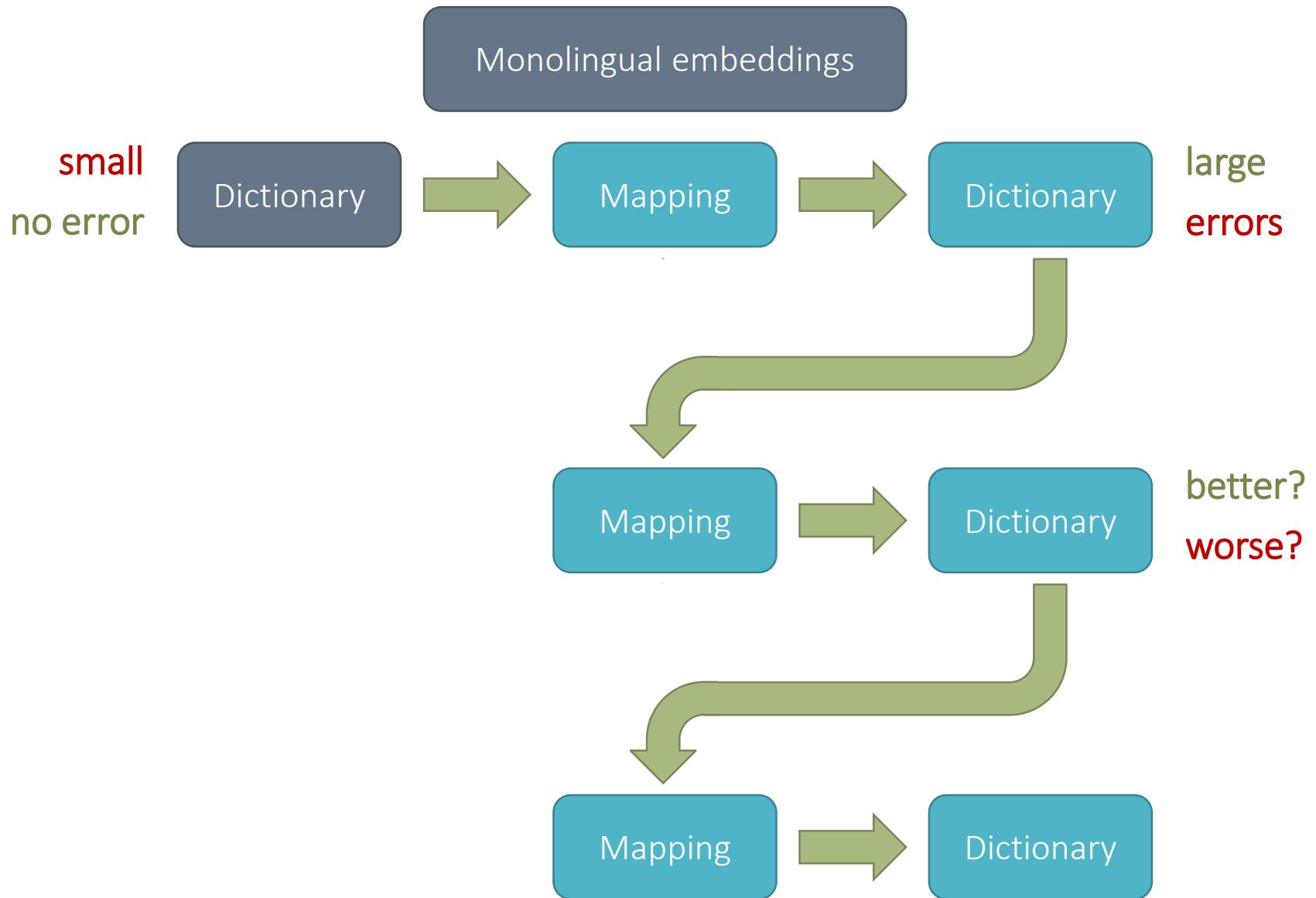
Why does it work?



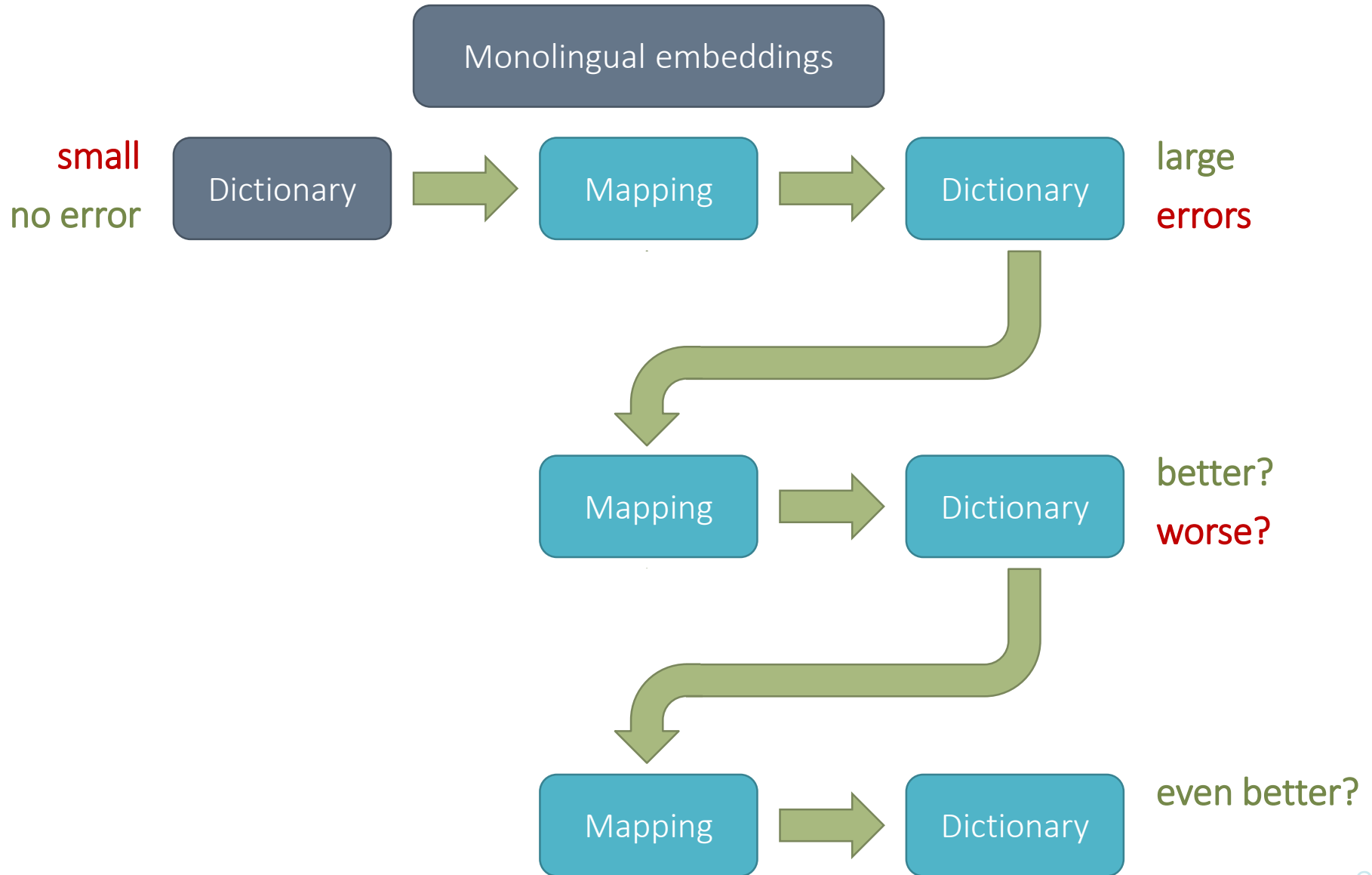
Why does it work?



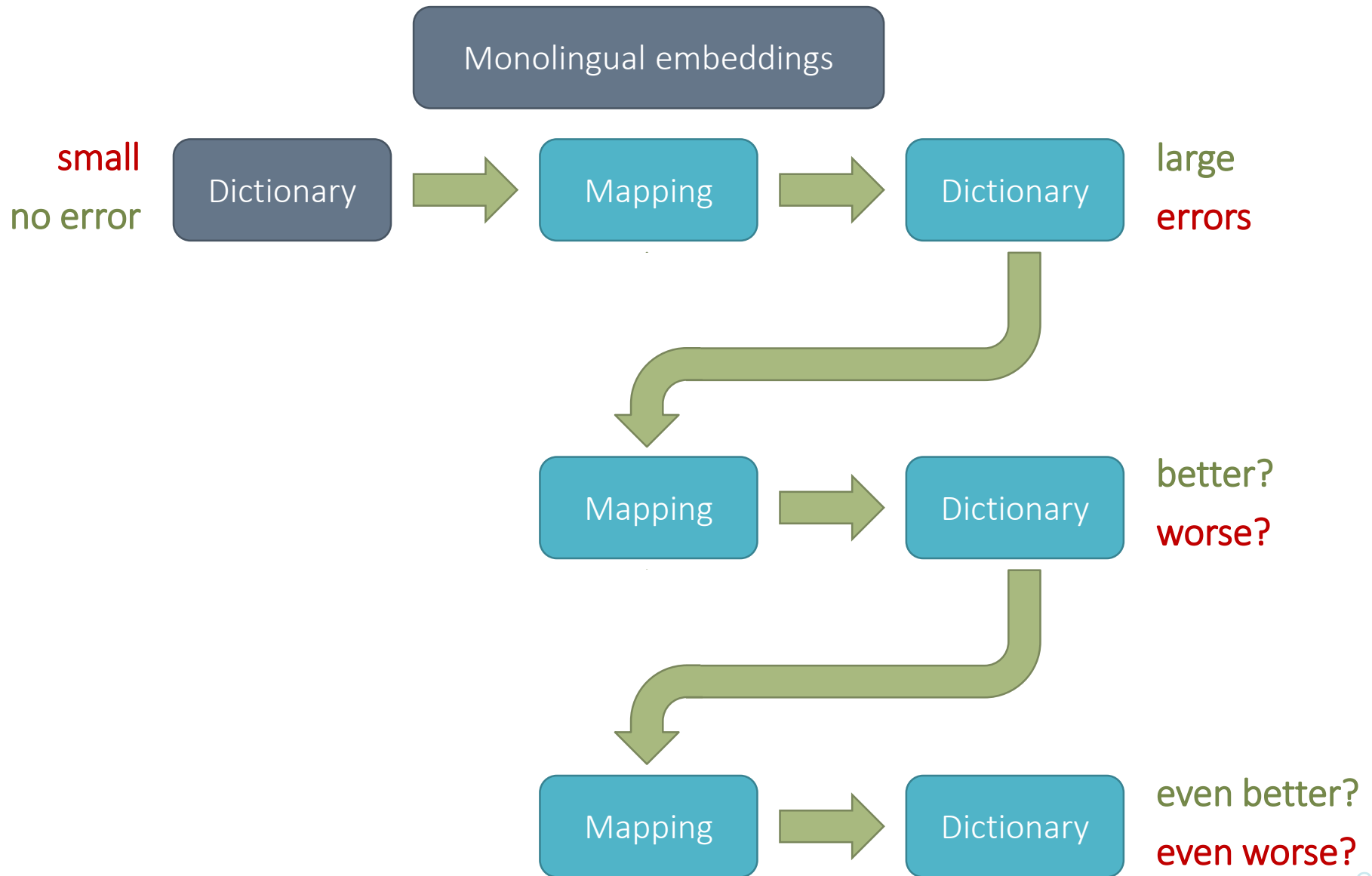
Why does it work?



Why does it work?



Why does it work?



Why does it work?

Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$

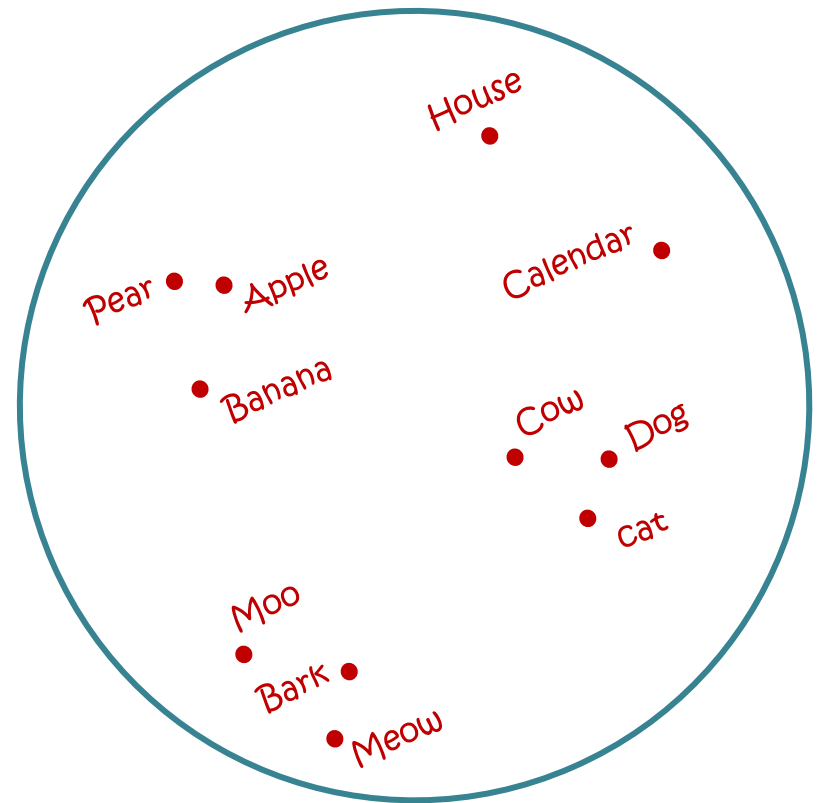
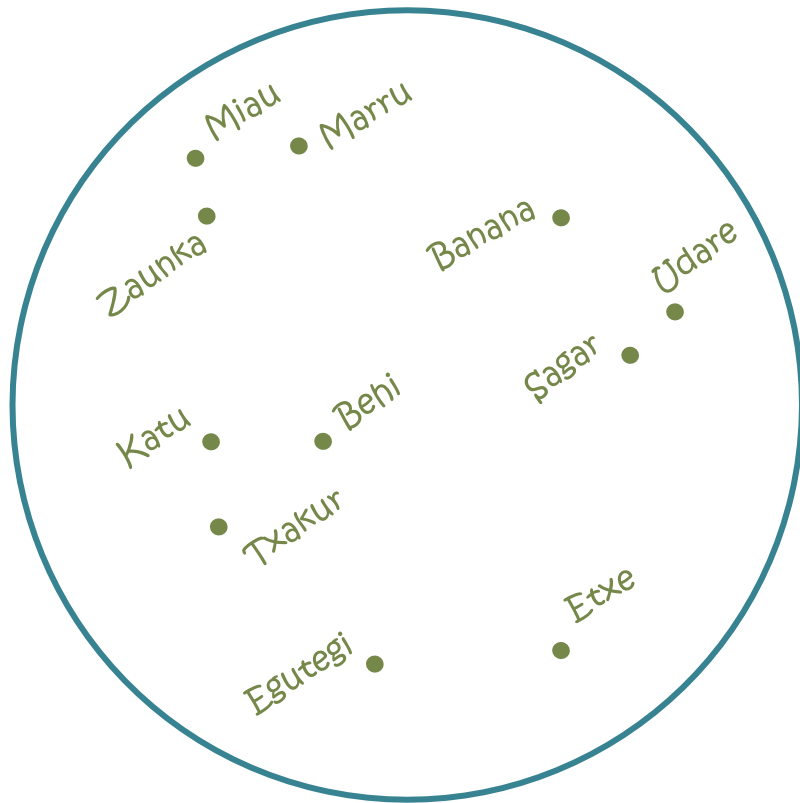
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$

Independent from seed dictionary!

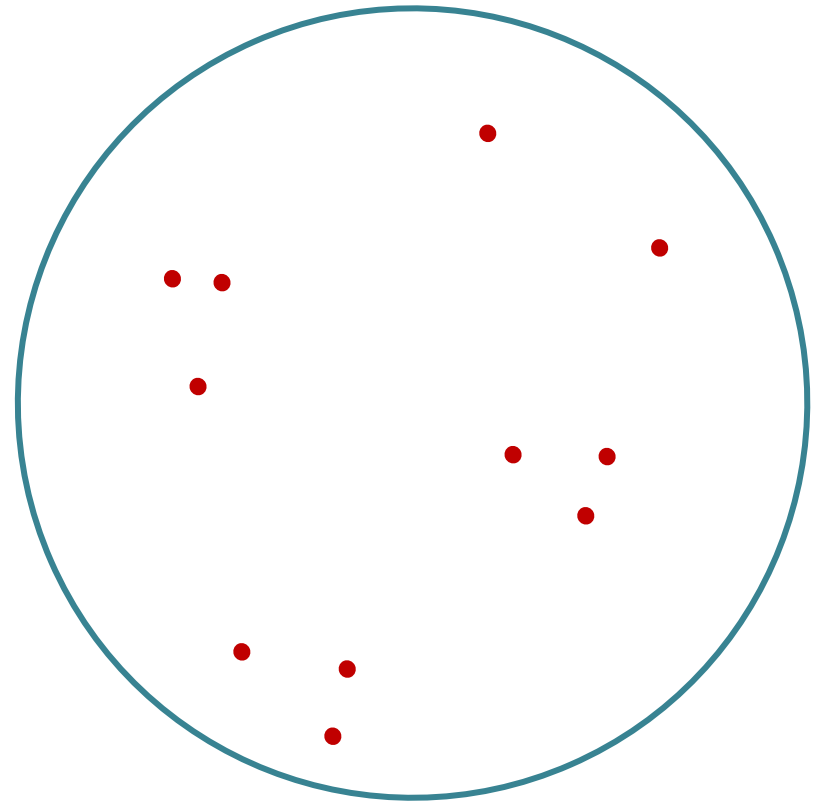
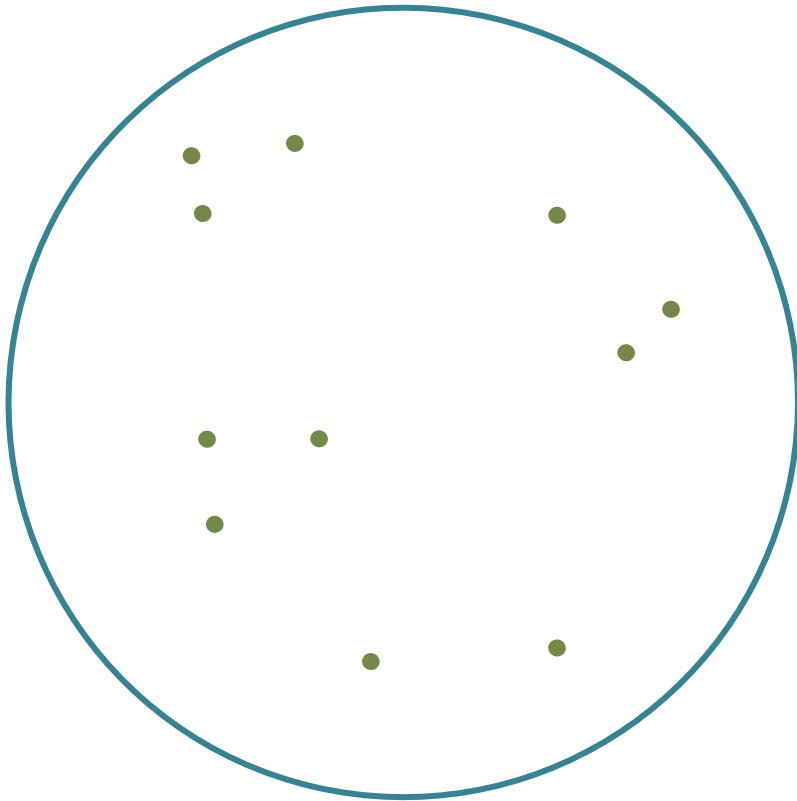
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



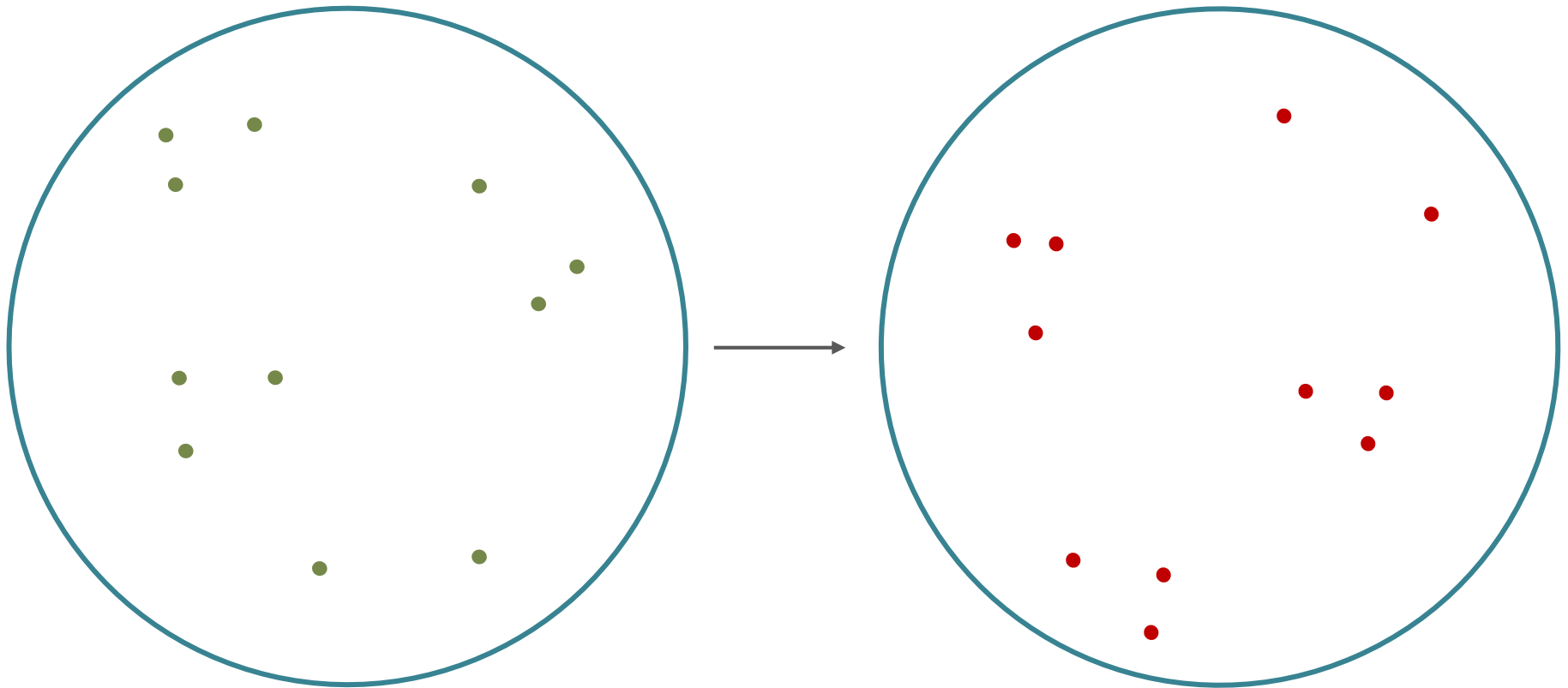
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



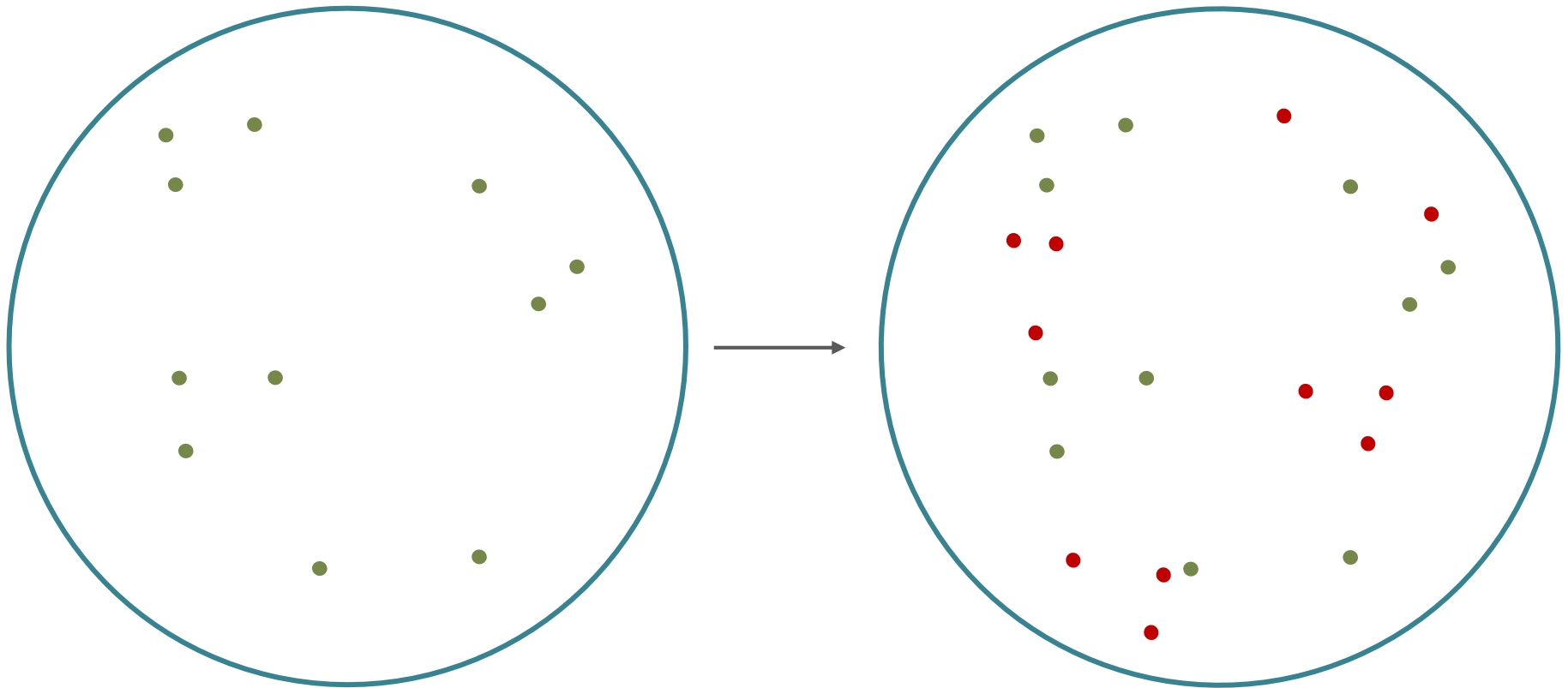
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$



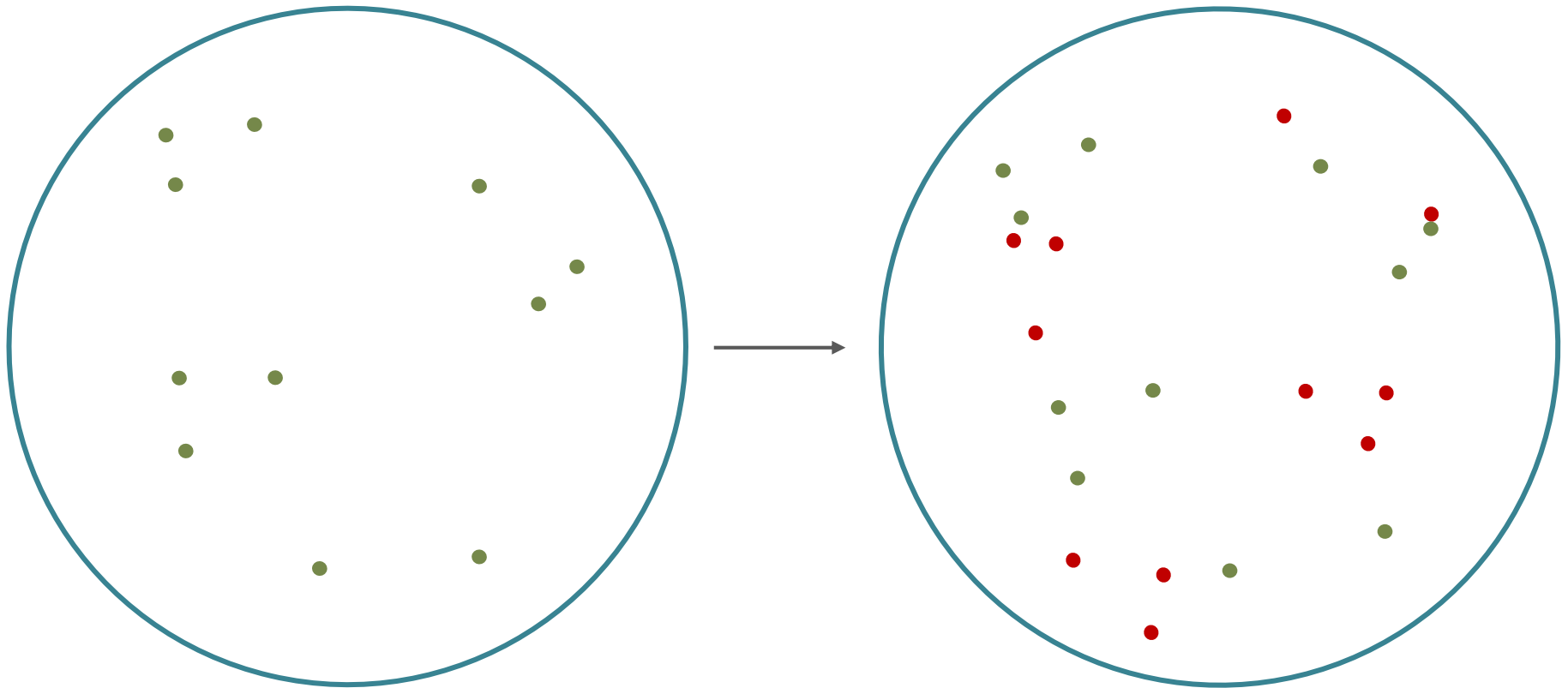
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



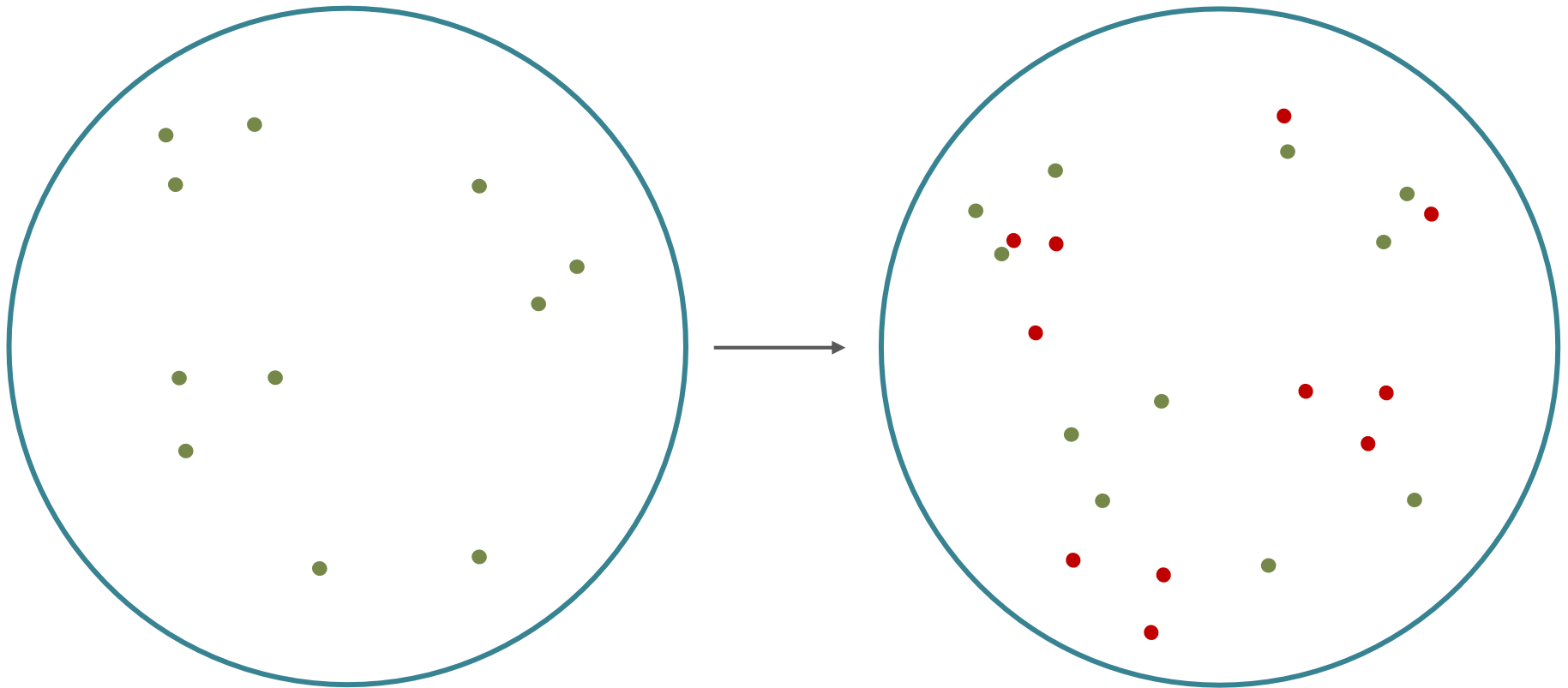
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$



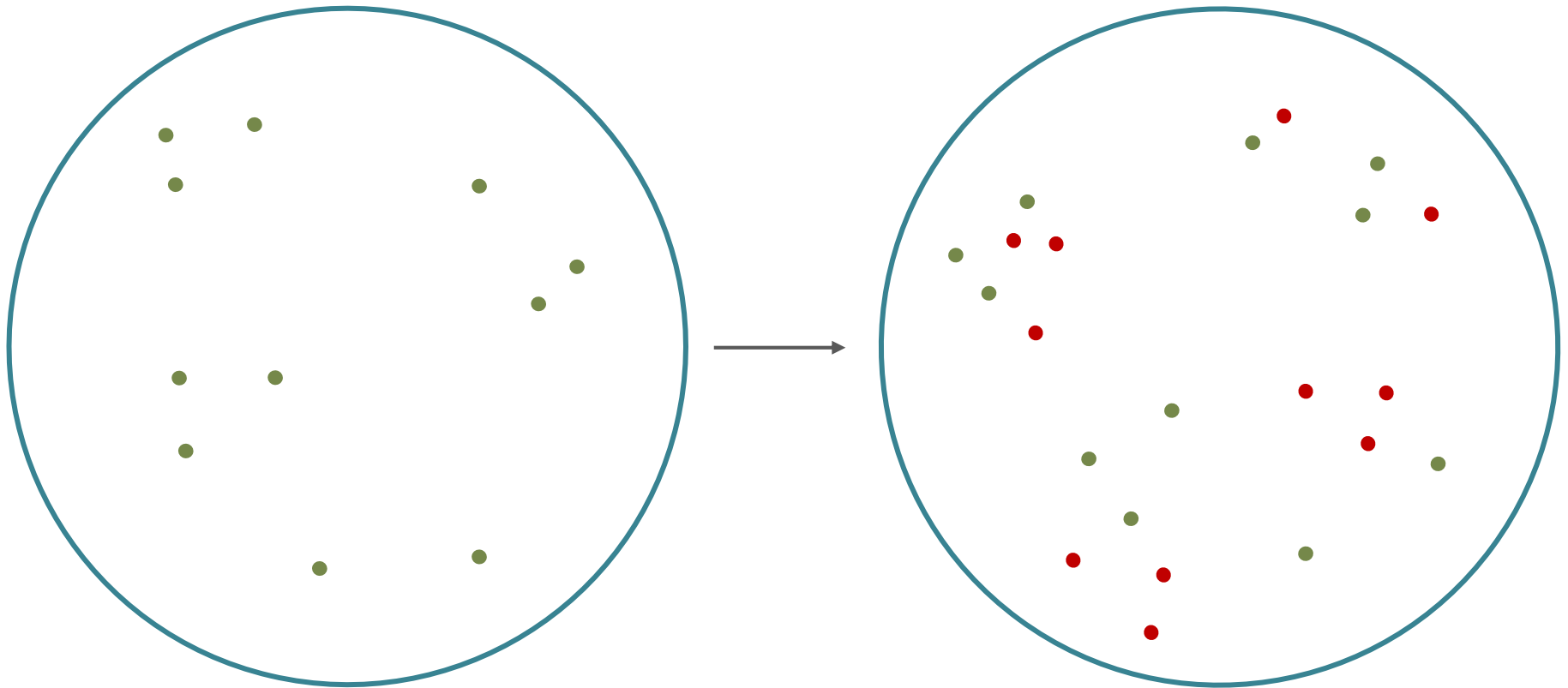
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



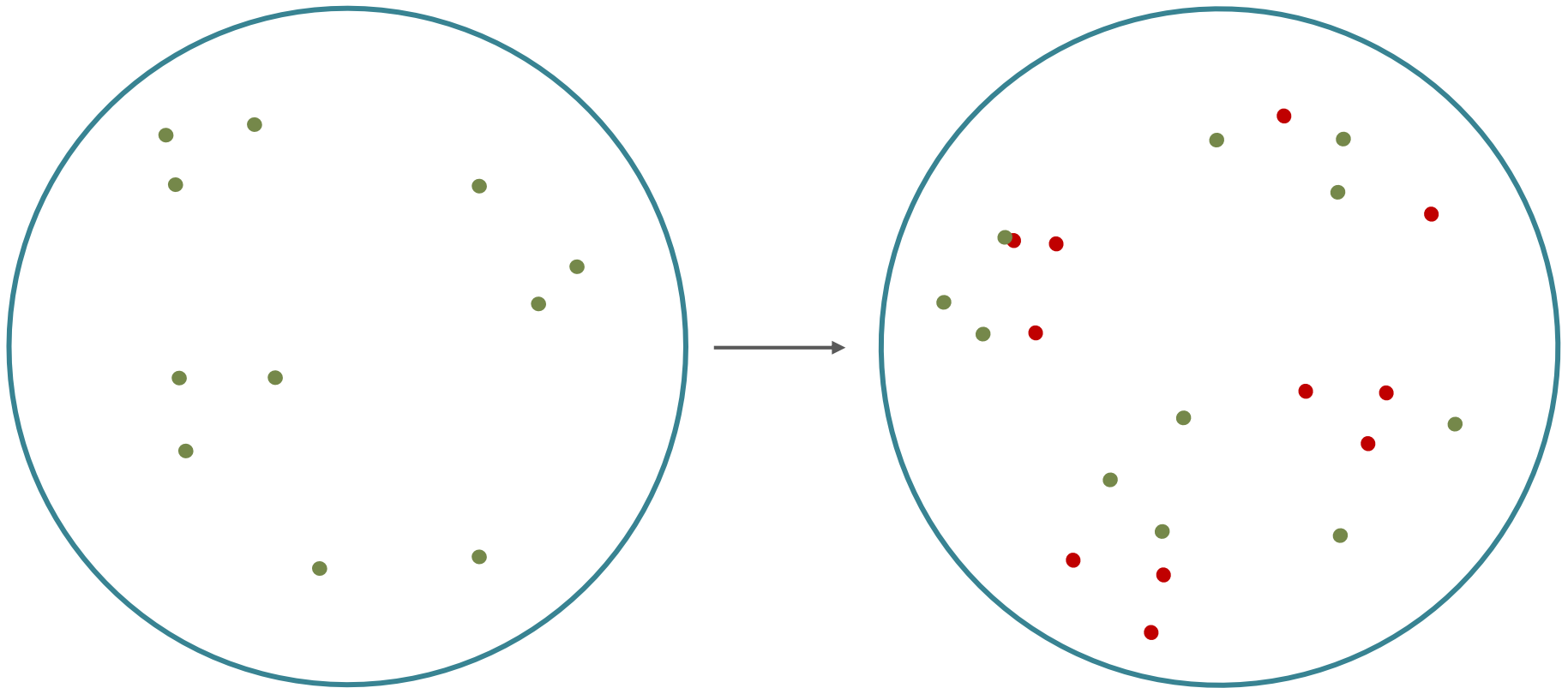
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



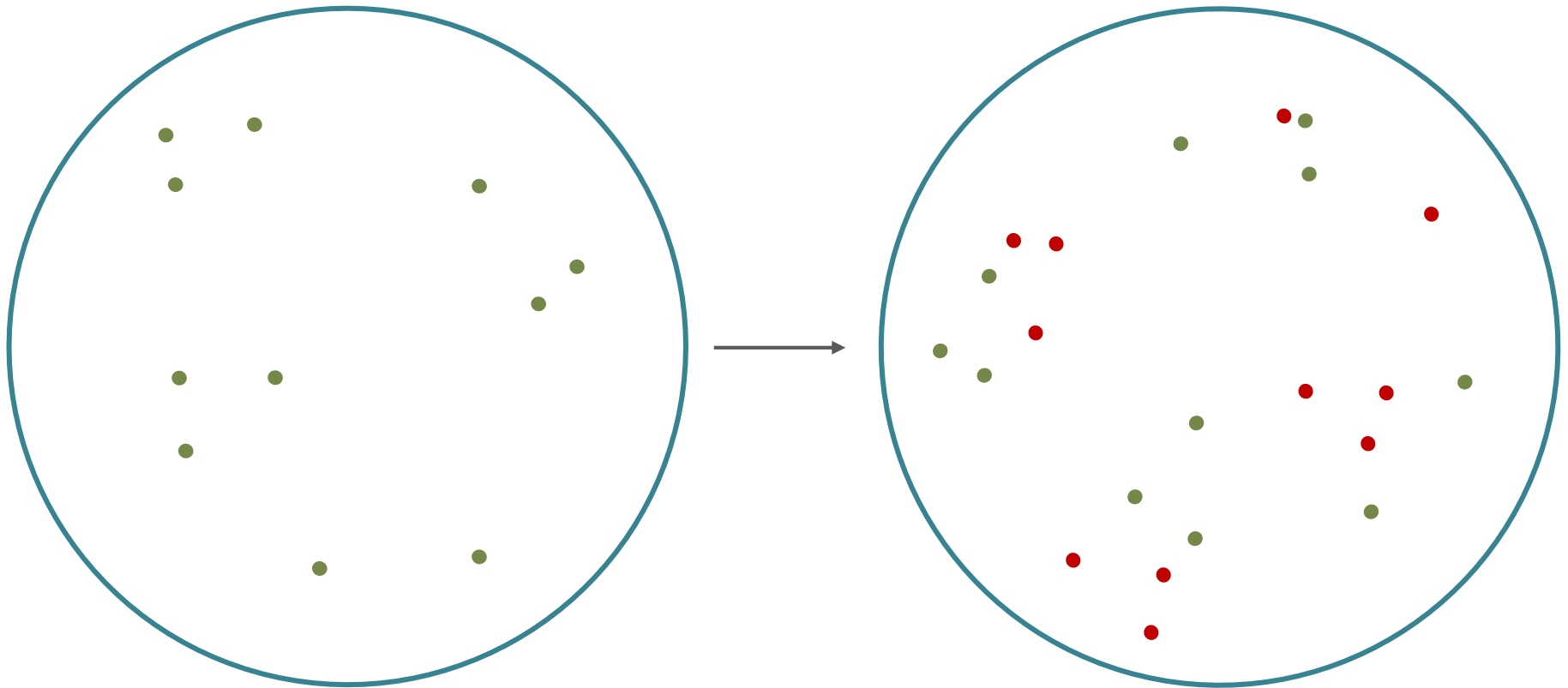
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



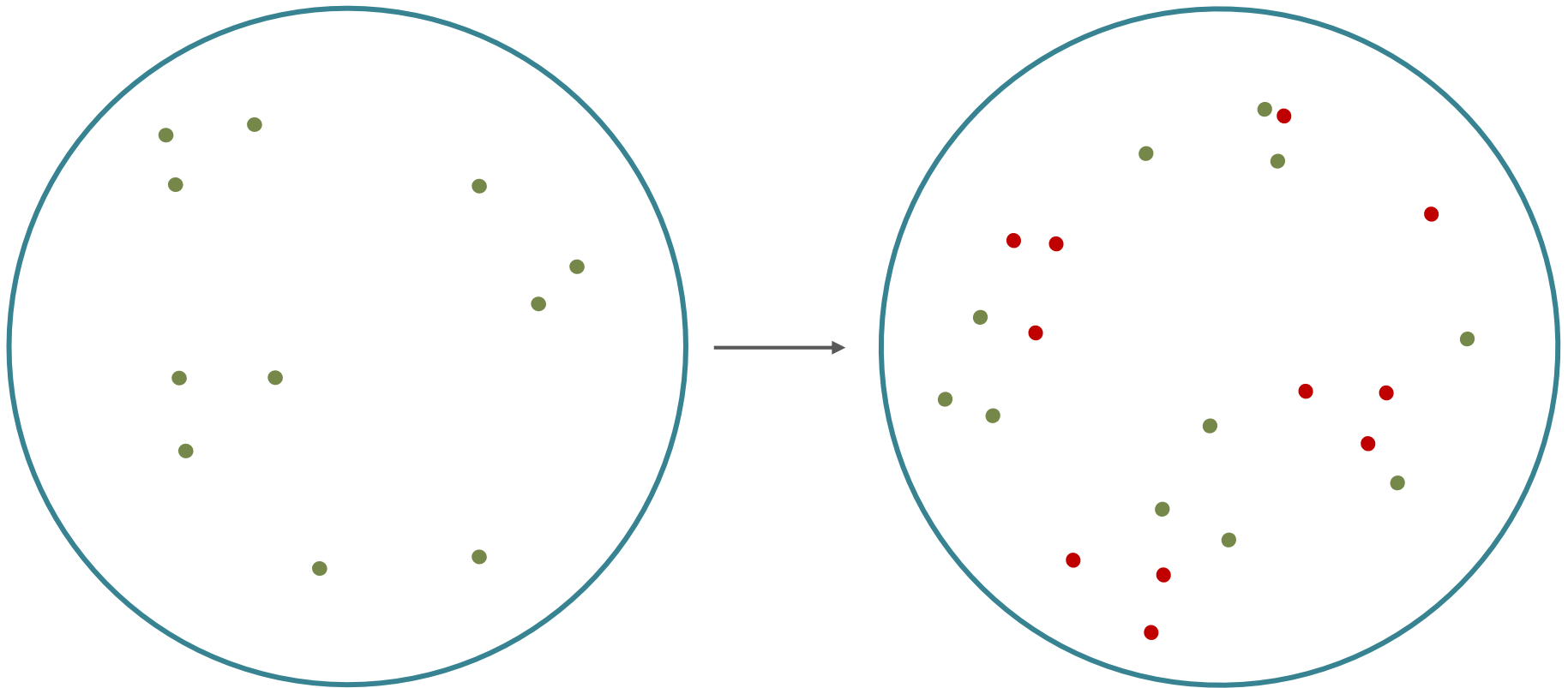
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



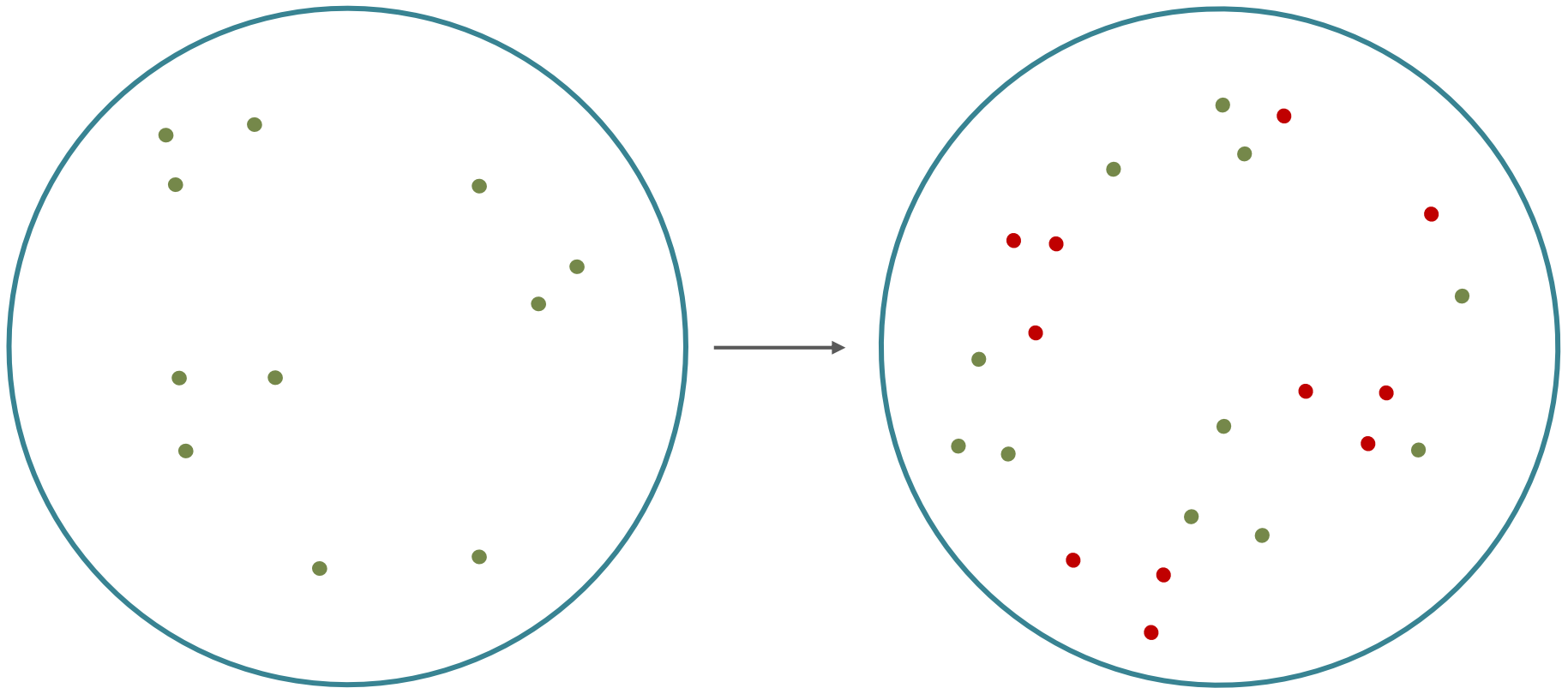
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



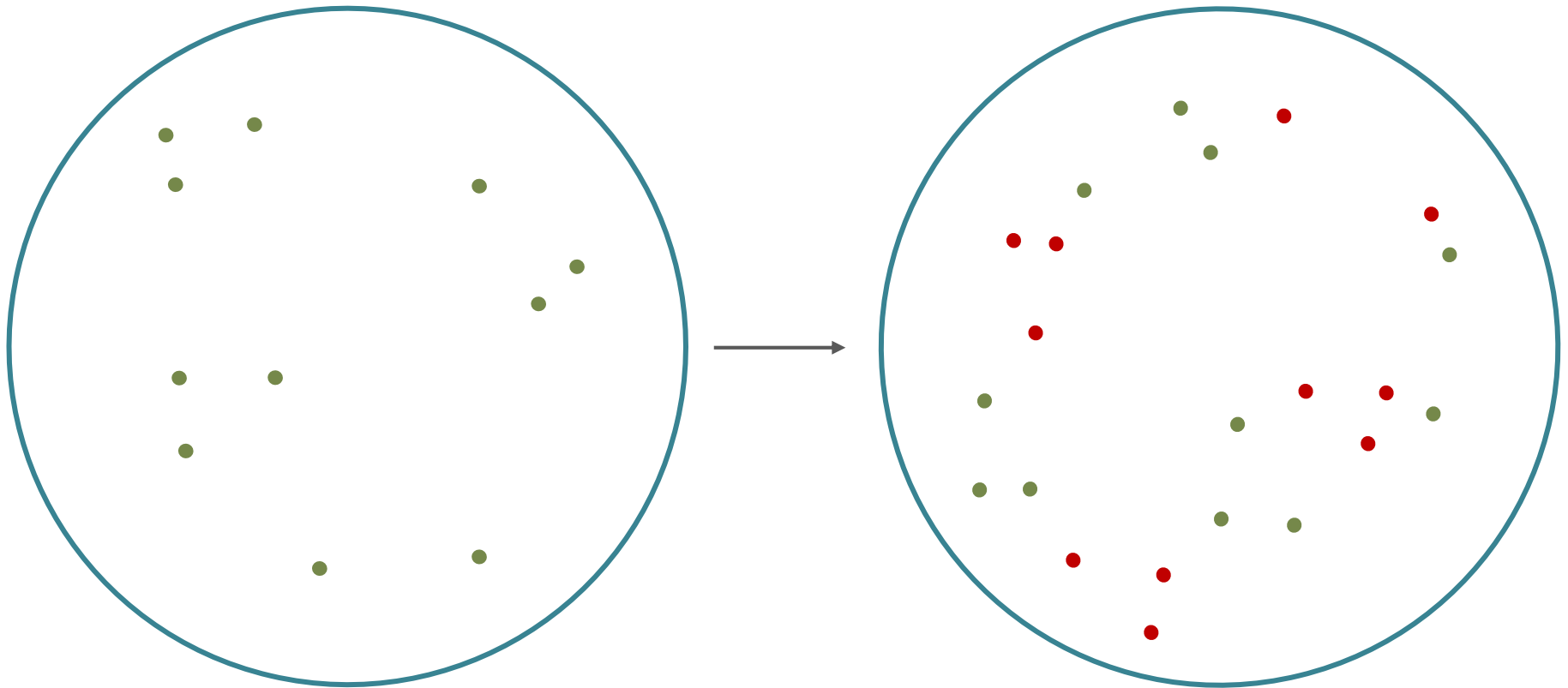
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



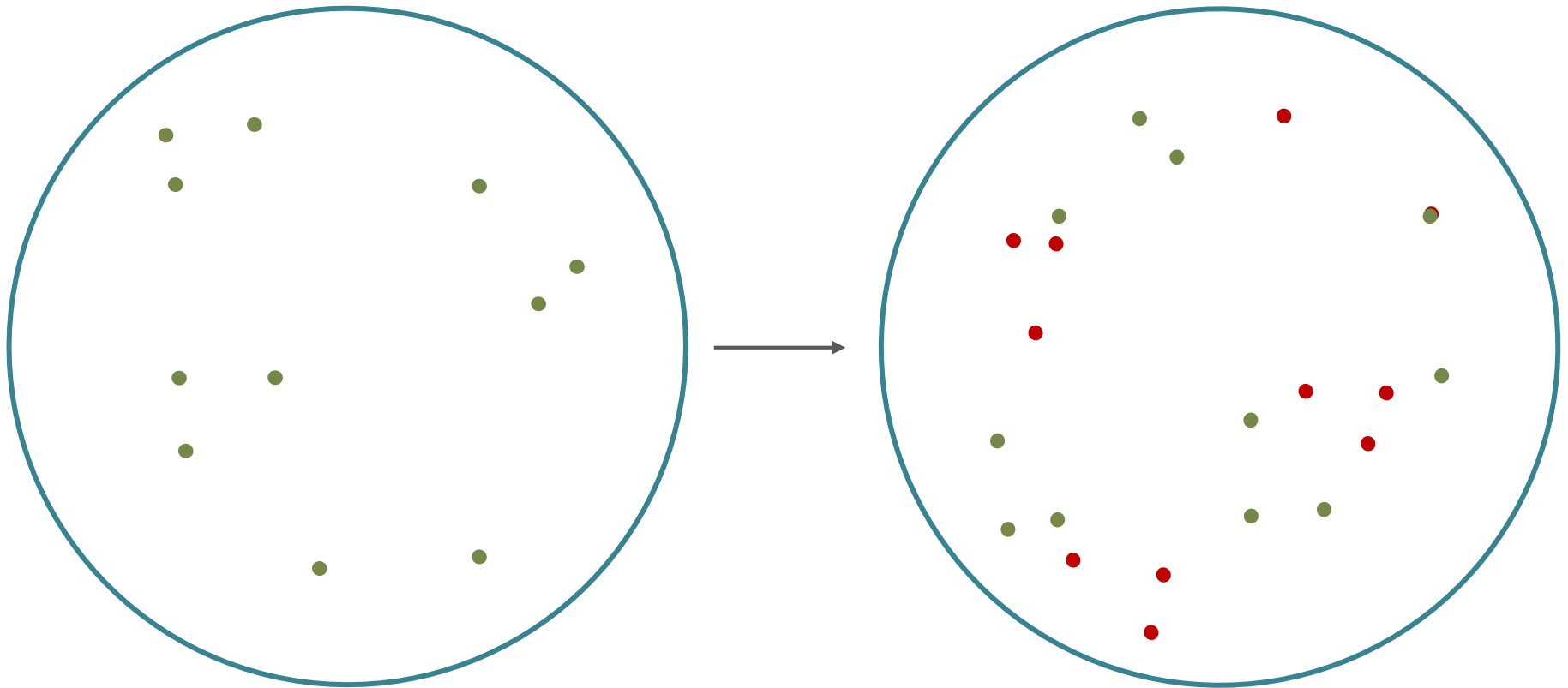
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



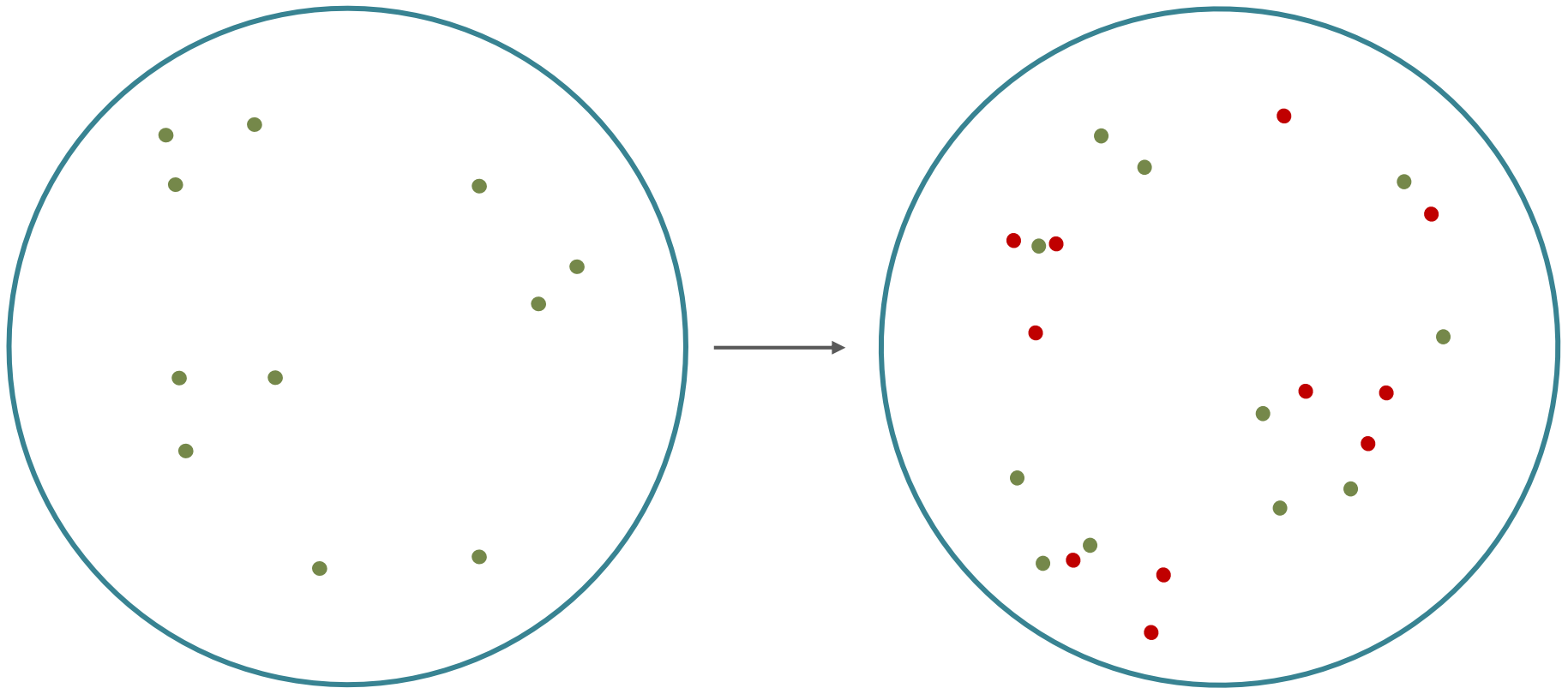
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



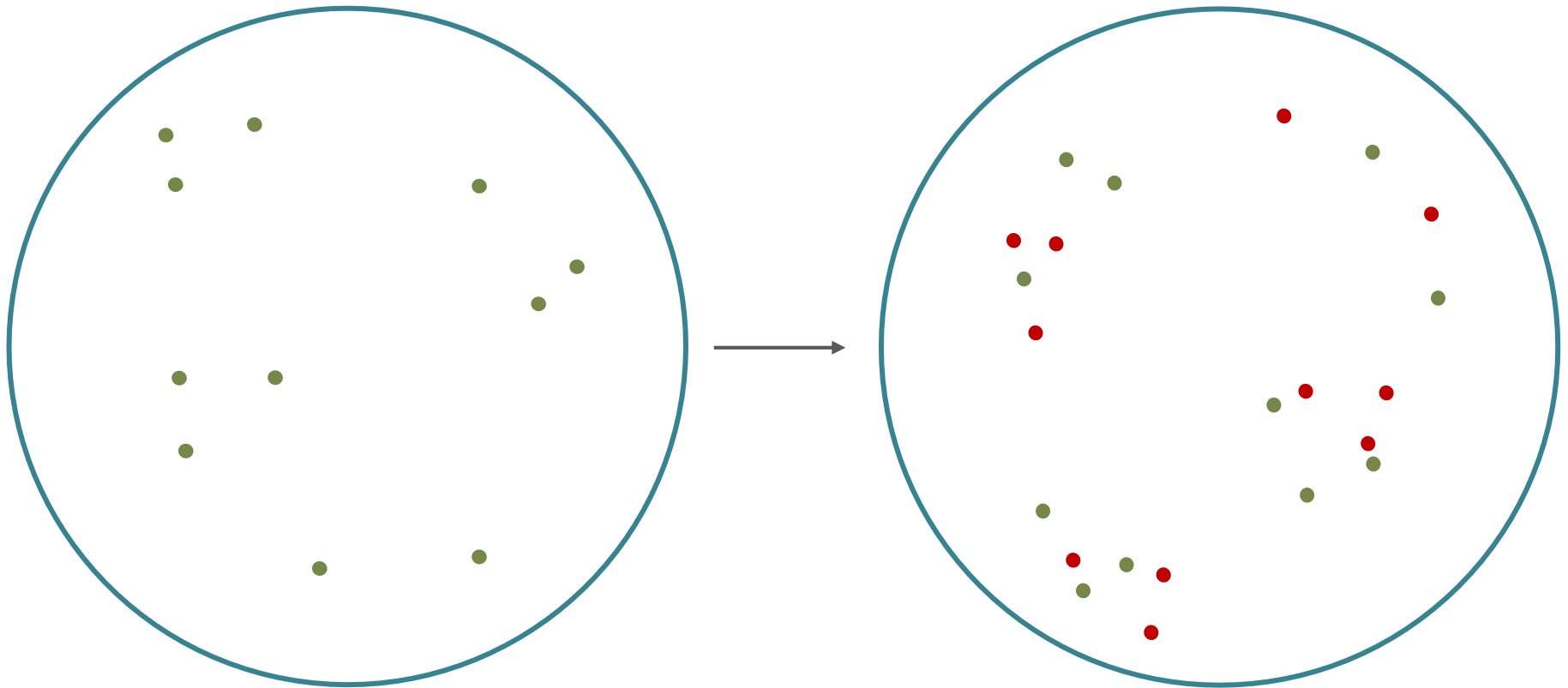
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



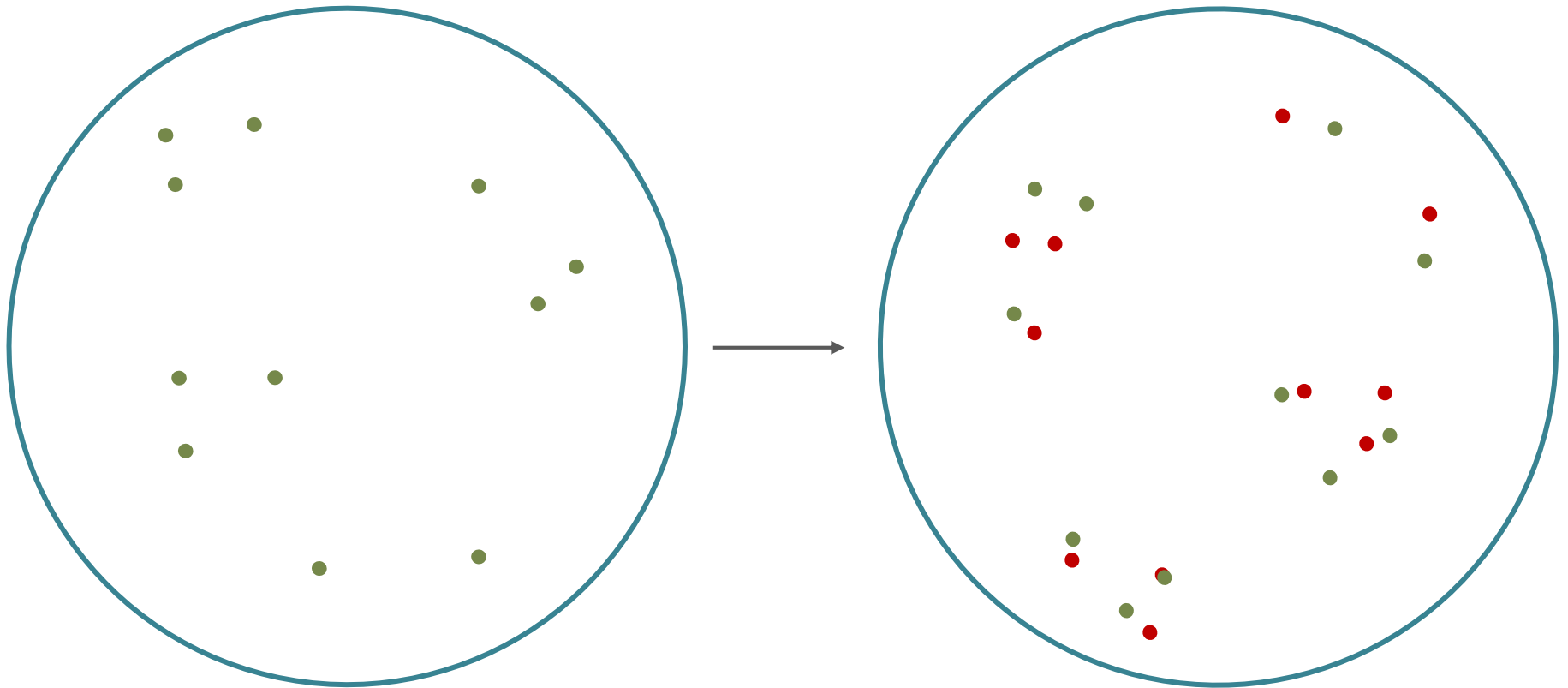
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$



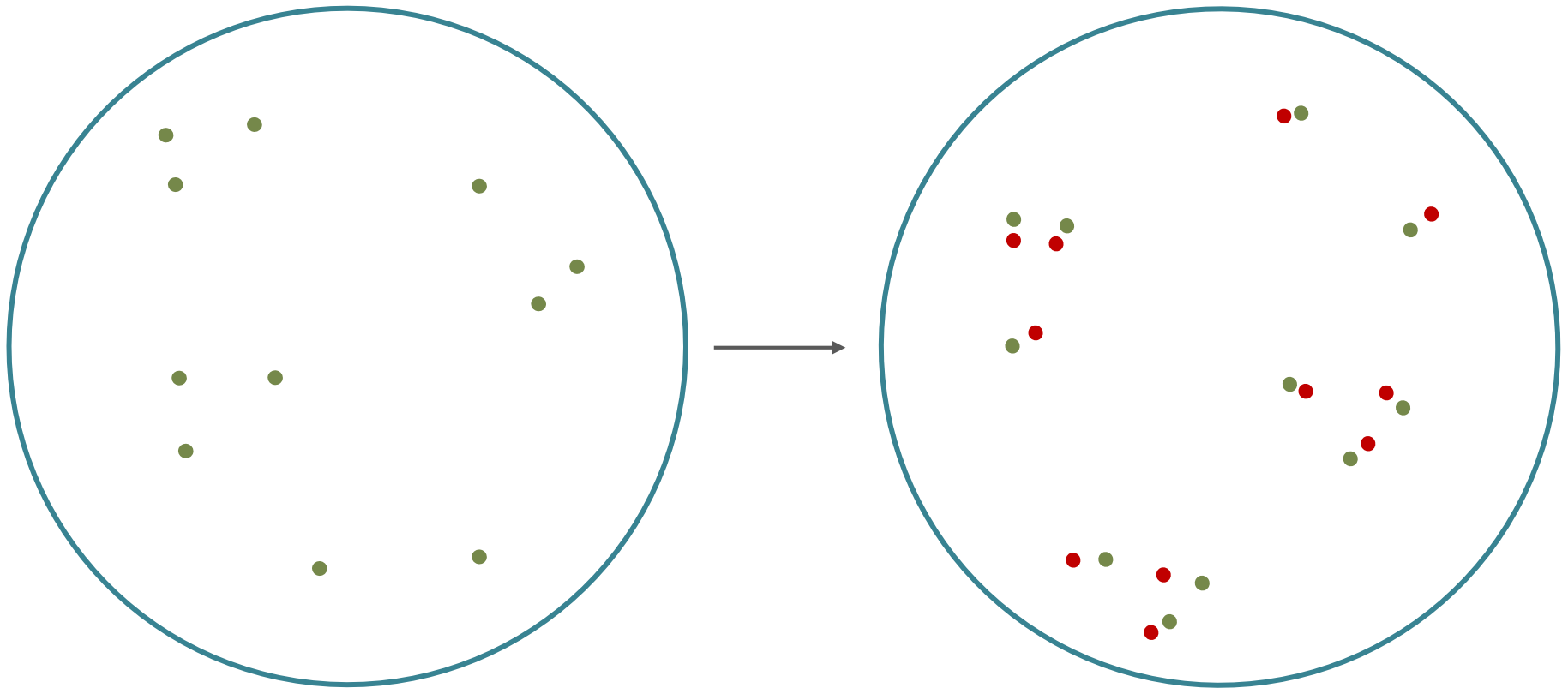
Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $WW^T = W^T W = I$



Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i^*} W) \cdot Z_{j^*} \quad \text{s.t. } WW^T = W^T W = I$

Independent from seed dictionary!

Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$

Independent from seed dictionary!

So why do we need a seed dictionary?

Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*} \quad \text{s.t. } WW^T = W^T W = I$

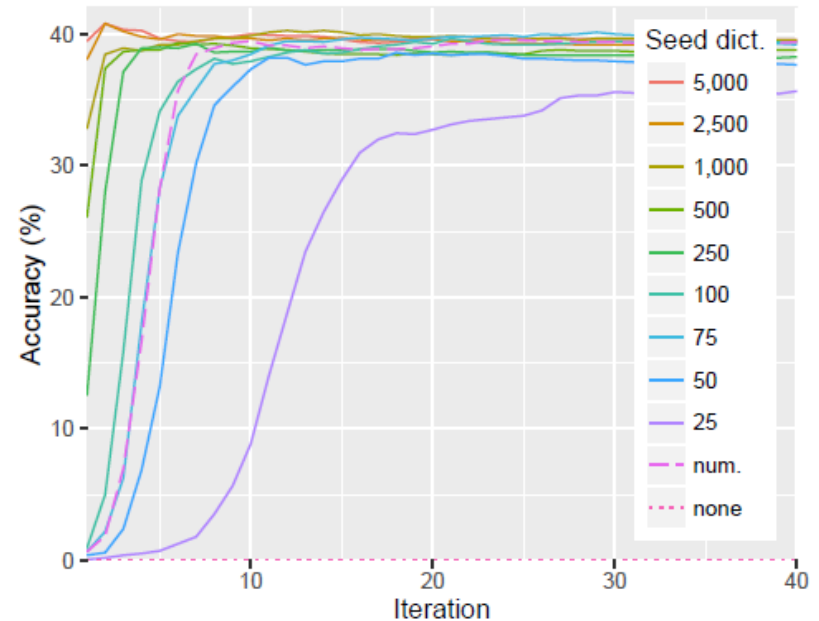
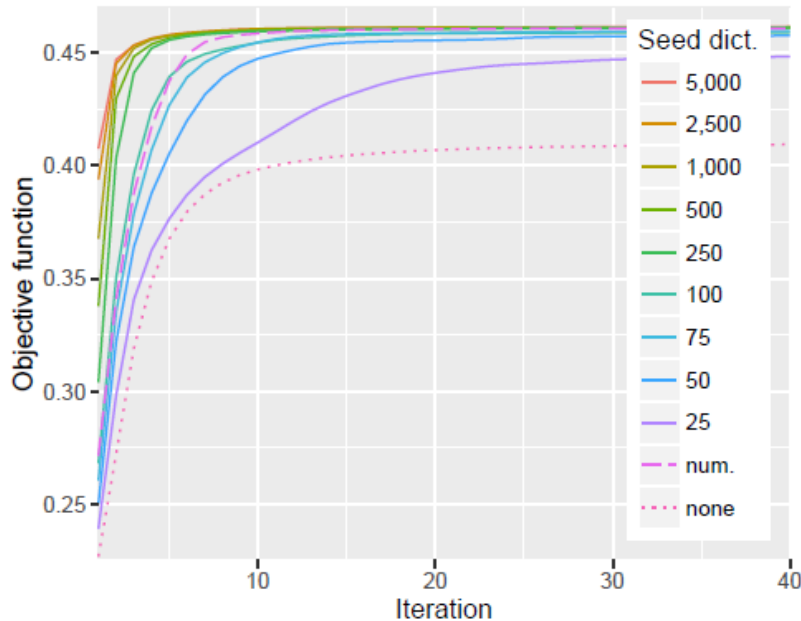
Independent from seed dictionary!

So why do we need a seed dictionary?

Avoid poor local optima!

Why does it work?

Implicit objective: $W^* = \arg \max_W \sum_i \max_j (X_{i*} W) \cdot Z_{j*}$ s.t. $W W^T = W^T W = I$



Conclusions

Conclusions

- Simple self-learning method to train bilingual embedding mappings

Conclusions

- Simple self-learning method to train bilingual embedding mappings
- High quality results with almost no supervision (25 words, numerals)

Conclusions

- Simple self-learning method to train bilingual embedding mappings
- High quality results with almost no supervision (25 words, numerals)
- Implicit optimization objective independent from seed dictionary

Conclusions

- Simple self-learning method to train bilingual embedding mappings
- High quality results with almost no supervision (25 words, numerals)
- Implicit optimization objective independent from seed dictionary
- Seed dictionary necessary to avoid poor local optima

Conclusions

- Simple self-learning method to train bilingual embedding mappings
- High quality results with almost no supervision (25 words, numerals)
- Implicit optimization objective independent from seed dictionary
- Seed dictionary necessary to avoid poor local optima
- Future work: fully unsupervised training

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
```

```
>
```

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 --self_learning --numerals`

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git  
> python3 vecmap/map_embeddings.py --self_learning --numerals  
SRC_INPUT.EMB TRG_INPUT.EMB
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git  
> python3 vecmap/map_embeddings.py --self_learning --numerals  
SRC_INPUT.EMB TRG_INPUT.EMB SRC_OUTPUT.EMB TRG_OUTPUT.EMB
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --self_learning --numerals
SRC_INPUT.EMB TRG_INPUT.EMB SRC_OUTPUT.EMB TRG_OUTPUT.EMB
>
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --self_learning --numerals
SRC_INPUT.EMB TRG_INPUT.EMB SRC_OUTPUT.EMB TRG_OUTPUT.EMB
> vecmap/reproduce_ac12017.sh
```

One more thing...

```
> git clone https://github.com/artetxem/vecmap.git
> python3 vecmap/map_embeddings.py --self_learning --numerals
SRC_INPUT.EMB TRG_INPUT.EMB SRC_OUTPUT.EMB TRG_OUTPUT.EMB
> vecmap/reproduce_acl2017.sh
```

ENGLISH-ITALIAN

5,000 WORD DICTIONARY

| | | | |
|--------------------------|--|---------------------|----------------|
| - Mikolov et al. (2013a) | | Translation: 34.93% | MWS353: 62.66% |
| - Xing et al. (2015) | | Translation: 36.87% | MWS353: 61.41% |
| - Zhang et al. (2016) | | Translation: 36.73% | MWS353: 61.62% |
| - Artetxe et al. (2016) | | Translation: 39.27% | MWS353: 61.74% |
| - Proposed method | | Translation: 39.67% | MWS353: 62.35% |

25 WORD DICTIONARY

| | | | |
|--------------------------|--|---------------------|----------------|
| - Mikolov et al. (2013a) | | Translation: 0.00% | MWS353: -6.42% |
| - Xing et al. (2015) | | Translation: 0.00% | MWS353: 19.49% |
| - Zhang et al. (2016) | | Translation: 0.07% | MWS353: 15.52% |
| - Artetxe et al. (2016) | | Translation: 0.07% | MWS353: 17.45% |
| - Proposed method | | Translation: 37.27% | MWS353: 62.64% |

NUMERAL DICTIONARY

| | | | |
|--------------------------|--|---------------------|----------------|
| - Mikolov et al. (2013a) | | Translation: 0.00% | MWS353: 28.75% |
| - Xing et al. (2015) | | Translation: 0.13% | MWS353: 27.75% |
| - Zhang et al. (2016) | | Translation: 0.27% | MWS353: 27.38% |
| - Artetxe et al. (2016) | | Translation: 0.40% | MWS353: 24.85% |
| - Proposed method | | Translation: 39.40% | MWS353: 62.82% |

Thank you!

<https://github.com/artetxem/vecmap>