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

Basque
- Monolingual corpus
- no cross-lingual signal

English
- Monolingual corpus
Introduction

Monolingual corpus

Basque

no cross-lingual signal

Monolingual corpus

English

cross-lingual embeddings
Introduction

Monolingual corpus

Basque

no cross-lingual signal

English

Monolingual corpus

cross-lingual embeddings
Introduction

- Unsupervised translation

Monolingual corpus

Basque

no cross-lingual signal

English

Monolingual corpus

cross-lingual embeddings
Introduction

- Unsupervised translation
- Cross-lingual transfer learning

Monolingual corpus

Basque

no cross-lingual signal

Monolingual corpus

English

cross-lingual embeddings
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

- Unsupervised translation
- Cross-lingual transfer 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

Materials
- Basque: Monolingual corpus
- English: Monolingual corpus
- No cross-lingual signal

- Unsupervised translation
- Cross-lingual transfer learning

Cross-lingual embeddings
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
- Cross-lingual transfer learning
- Unsupervised translation

Monolingual corpus

Basque
- No cross-lingual signal

English
- Monolingual corpus

Cross-lingual embeddings
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

Basque
Monolingual corpus
no cross-lingual signal

English
Monolingual corpus

- Unsupervised translation
- Cross-lingual transfer learning

cross-lingual embeddings
Cross-lingual embedding mappings
Cross-lingual embedding mappings
Cross-lingual embedding mappings
Cross-lingual embedding mappings

Basque

Training dictionary

English

X

Z

Miau
Marru
Zaunka
Katu
Txakur
Egutegi
Etxe
Banana
Čagar
Udare

House
Calendar
Cow
Dog
Meow
Bark
Moo
Banana
Apple
Pear

cat
Cross-lingual embedding mappings

Basque

Training dictionary

English

Txakur
Sagar
:
Egutegi

Dog
Apple
:
Calendar
Cross-lingual embedding mappings

Training dictionary

Basque

Txakur
Sagar
Egutegi

English

Dog
Apple
Calendar

W
Cross-lingual embedding mappings

Basque

Training dictionary

English

Txakur
Sagar
Egutegi

Dog
Apple
Calendar
Cross-lingual embedding mappings

$$
\begin{align*}
X_1, & \cdots, X_n, \\
Z_1, & \cdots, Z_n
\end{align*}
$$

$$
W \approx Z_1, \cdots, Z_n
$$

Basque
Suunka
Miau
Marru
Zaunka
Katu
Behi
Sagar
Udare
Txakur
Egutegi
Etxe

Training dictionary

English
House
Etxe
Calendar
Cow
Dog
Cat
Moo
Marru
Bark
Miau
Zaunka
Meow

$$
\begin{bmatrix}
X_1,* \\
X_2,* \\
\vdots \\
X_n,*
\end{bmatrix}
$$

$$
\begin{bmatrix}
Z_1,* \\
Z_2,* \\
\vdots \\
Z_n,*
\end{bmatrix}
$$

Txakur
Sagar
Egutegi
Cross-lingual embedding mappings

\[
X_1, * X_2, * \ldots X_n, * \\
\approx \approx \\
Z_1, * Z_2, * \ldots Z_n, *
\]

Training dictionary

Basque

Egutegi
Miau
Marru
Katu
Txakur
Etxe
Bark
Miau
Meow

English

Dog
Apple
Calendar

X

Z

XW
Cross-lingual embedding mappings

\[
X_1, * \quad X_2, * \quad \vdots \quad X_n, *
\]

\[
W \approx Z_1, * \quad Z_2, * \quad \vdots \quad Z_n, *
\]

\[
\text{arg min}_{W \in O(n)} \sum_i \| X_iW - Z_j^* \|^2
\]

Basque

English

Txakur
Sagar
Egutegi

Dog
Apple
Calendar
Cross-lingual embedding mappings

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

\[
\begin{align*}
\text{Txakur} & \quad \left[ X_{1,*} \right] \\
\text{Sagar} & \quad \left[ X_{2,*} \right] \\
\vdots & \quad \vdots \\
\text{Egutegi} & \quad \left[ X_{n,*} \right]
\end{align*}
\]

\[
\begin{align*}
\text{Dog} & \quad \left[ Z_{1,*} \right] \\
\text{Apple} & \quad \left[ Z_{2,*} \right] \\
\vdots & \quad \vdots \\
\text{Calendar} & \quad \left[ Z_{n,*} \right]
\end{align*}
\]
Cross-lingual embedding mappings

\[
\text{arg min}_{W \in O(n)} \sum_i \|X_iW - Z_j\|^2
\]

\[
\begin{pmatrix}
X_1,* \\
X_2,* \\
\vdots \\
X_n,*
\end{pmatrix} [W] \approx
\begin{pmatrix}
Z_1,* \\
Z_2,* \\
\vdots \\
Z_n,*
\end{pmatrix}
\]

Basque

- Mjau
- Marru
- Zaunka
- Katu
- Behi
- Sagar
- Udare
- Txakur
- Egutegi
- Etxe

English

- House
- Etxe
- Dog
- Apple
- Calendar
- Cow
- Meow
- Bark
- Mjau
- Zaunka
- Katu
- Cat

Txakur

Sagar

Egutegi

\[W\]
Cross-lingual embedding mappings

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

\[
\begin{bmatrix}
X_{1,*} \\
X_{2,*} \\
\vdots \\
X_{n,*}
\end{bmatrix}
\begin{bmatrix}
W
\end{bmatrix} \approx
\begin{bmatrix}
Z_{1,*} \\
Z_{2,*} \\
\vdots \\
Z_{n,*}
\end{bmatrix}
\]

Basque

- Mjau
- Marru
- Norki
- Behi
- Sagar
- Udare
- Egutegi
- Txakur
- Etxe

English

- House
- Etxe
- Pear
- Udare
- Sagar
- Calendar
- Cow
- Dog
- Bark
- Mjau
- Zunka
- Cat
- Meow

Txakur
Sagar
Egutegi
Cross-lingual embedding mappings

\[ X_1, * X_2, * \ldots X_n, * W \approx Z_1, * Z_2, * \ldots Z_n, * \]

\[ \arg \min_W \sum_i \| X_i W - Z_i \| \]

Basque

- Mjau
- Marru
- Zaunka
- Katu
- Behi
- Sagar
- Udare
- Txakur
- Egutegi
- Etxe

English

- House
- Etxe
- Calendar
- Eggutegi
- Dog
- Cow
- Apple
- Pear
- Sagar
- Udare
- Behi
- Calander
- Moo
- Mjau
- Marru
- Bark
- Meow
- Zunka
- Ma
- Etxe

Txakur

Sagar

Egutegi

\[ \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_n,* \end{bmatrix} \begin{bmatrix} W \end{bmatrix} \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_n,* \end{bmatrix} \begin{bmatrix} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{bmatrix} \]
Cross-lingual embedding mappings

Basque
- Zaunka
- Katu
- Txakur
- Egutegi
- X
- Marru
- Behi
- Sagar
- Udare
- Etxe

English
- Z
- House
- Calendar
- Pear
- Apple
- Cow
- Dog
- cat
- Bark
- Moo
- Meow
- Banana
- Ban
Cross-lingual embedding mappings
Cross-lingual embedding mappings

$X \xrightarrow{W} Z$
Cross-lingual embedding mappings

$W \rightarrow X \rightarrow Z \rightarrow XW$
Cross-lingual embedding mappings

\[ W \]

![Diagram showing cross-lingual embedding mappings](image-url)
Cross-lingual embedding mappings
Cross-lingual embedding mappings
Cross-lingual embedding mappings
Cross-lingual embedding mappings

$X \rightarrow W \rightarrow Z \rightarrow XW$
Cross-lingual embedding mappings
Cross-lingual embedding mappings
Cross-lingual embedding mappings

\[
W \quad \rightarrow \quad W \cdot X
\]
Cross-lingual embedding mappings
Cross-lingual embedding mappings

\[ W \]
Cross-lingual embedding mappings

\[ W \]

\[ X \rightarrow W \rightarrow Z, XW \]
Cross-lingual embedding mappings
Cross-lingual embedding mappings

\[ W X Z XW \]
Cross-lingual embedding mappings

\[ W^* = \arg \min_{W \in O(n)} \sum_i \min_j \|X_iW - Z_j^*\|^2 \]
Cross-lingual embedding mappings

\[ W^* = \arg \min_{W \in \Theta(n)} \sum_i \min_j \|X_i W - Z_j^*\|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \theta(n)} \sum_i \min_j \| X_{i*}W - Z_{j*} \|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \mathcal{O}(n)} \sum_{i} \min_{j} \|X_{i*}W - Z_{j*}\|^2 \]
\[ W^* = \arg \min_{W \in \Omega(n)} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2 \]
\begin{equation}
W^* = \arg \min_{W \in \theta(n)} \sum_i \min_j \|X_i^* W - Z_j^*\|^2
\end{equation}
\[ W^* = \arg \min_{W \in \Theta(n)} \sum_i \min_j \|X_i W - Z_j^*\|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \mathcal{O}(n)} \sum_i \min_j \|X_i W - Z_j^*\|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \mathbb{O}(n)} \sum_i \min_j \|X_i^* W - Z_j^*\|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \Theta(n)} \sum_i \min_j \| X_i W - Z_j \|^2 \]

- 25 word pairs
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min \sum_i \min_j \|X_i W - Z_j\|^2 \]

- 25 word pairs
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in O(n)} \sum_i \min_j \|X_i W - Z_j^*\|^2 \]

- 25 word pairs
- Numeral list
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \Omega(n)} \sum_i \min_j \|X_i W - Z_j \|^2 \]

- 25 word pairs
- Numeral list

Diagram:
- Dictionary
- Mapping
- Dictionary
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in O(n)} \sum_i \min_j \|X_i W - Z_j^*\|^2 \]

- 25 word pairs
- Numeral list
- Random dict.
Artetxe et al. (ACL 2017)

\[ W^* = \arg\min_{W \in O(n)} \sum_i \min_j \|X_iW - Z_j\|^2 \]
Artetxe et al. (ACL 2017)

\[ W^* = \arg \min_{W \in \mathcal{O}(n)} \sum_i \min_j \| X_{i*}W - Z_{j*} \|^2 \]

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

Dictionary → Mapping → Dictionary
Artetxe et al. (ACL 2017)

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

Diagram:
- Dictionary
  - Mapping
    - Dictionary
    - Dictionary
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

2) Robust self-learning
Proposed method

1) Fully unsupervised initialization

2) Robust self-learning

two

**Intra-lingual similarity distribution**
Proposed method

1) Fully unsupervised initialization

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

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

2) Robust self-learning

Intra-lingual similarity distribution
Proposed method

1) Fully unsupervised initialization

<table>
<thead>
<tr>
<th>English</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

**Intra-lingual similarity distribution**

two  due

2) Robust self-learning
Proposed method

1) Fully unsupervised initialization

Intra-lingual similarity distribution

2) Robust self-learning
Proposed method

1) Fully unsupervised initialization

2) Robust self-learning

Intra-lingual similarity distribution
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](image)

2) Robust self-learning
Proposed method

1) Fully unsupervised initialization

2) Robust self-learning

**Intra-lingual similarity distribution**
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

[Graph showing intra-lingual similarity distribution for English and Italian words: two, due (two), cane (dog).

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

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

Intra-lingual similarity distribution

2) Robust self-learning
   - Stochastic dictionary induction
   - Frequency-based vocabulary cutoff
Proposed method

1) Fully unsupervised initialization

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

Intra-lingual similarity distribution

2) Robust self-learning
- Stochastic dictionary induction
- Frequency-based vocabulary cutoff
- CSLS retrieval (Conneau et al., 2018)
Proposed method

1) Fully unsupervised initialization

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

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

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

Intra-lingual similarity distribution

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
Experiments

- Bilingual lexicon extraction

<table>
<thead>
<tr>
<th>Method</th>
<th>Zhang et al. (2017), $\lambda = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhang et al. (2017), $\lambda = 10$</td>
</tr>
</tbody>
</table>
Experiments

- Bilingual lexicon extraction

<table>
<thead>
<tr>
<th>Method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td></td>
</tr>
</tbody>
</table>
Experiments

- Bilingual lexicon extraction

<table>
<thead>
<tr>
<th>Method</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
</tr>
</tbody>
</table>
Experiments

• Bilingual lexicon extraction
• 10 runs for each method

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>
Experiments

• Bilingual lexicon extraction
• 10 runs for each method
  ⇒ Best/average accuracy

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>
Experiments

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

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>
Experiments

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

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>
Experiments

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

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), ( \lambda = 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2017), ( \lambda = 10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiments

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

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

Best accuracy (%)
Experiments

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

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), ( \lambda = 1 )</td>
<td>71.43</td>
<td>60.38</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), ( \lambda = 10 )</td>
<td>70.24</td>
<td>57.64</td>
<td>21.07</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>76.18</td>
<td>67.32</td>
<td>32.64</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>76.15</td>
<td>67.21</td>
<td>29.79</td>
</tr>
<tr>
<td>Proposed method</td>
<td>76.43</td>
<td>66.96</td>
<td>36.10</td>
</tr>
</tbody>
</table>

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)**

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>71.43</td>
<td>60.38</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>70.24</td>
<td>57.64</td>
<td>21.07</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>76.18</td>
<td><strong>67.32</strong></td>
<td>32.64</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>76.15</td>
<td>67.21</td>
<td>29.79</td>
</tr>
<tr>
<td>Proposed method</td>
<td><strong>76.43</strong></td>
<td>66.96</td>
<td><strong>36.10</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>71.43</td>
<td>60.38</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>70.24</td>
<td>57.64</td>
<td>21.07</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>76.18</td>
<td>67.32</td>
<td>32.64</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>76.15</td>
<td>67.21</td>
<td>29.79</td>
</tr>
<tr>
<td>Proposed method</td>
<td><strong>76.43</strong></td>
<td>66.96</td>
<td><strong>36.10</strong></td>
</tr>
</tbody>
</table>

*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)

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>68.18</td>
<td>56.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>66.37</td>
<td>52.60</td>
<td>17.95</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>75.82</td>
<td><strong>67.00</strong></td>
<td>14.34</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>75.81</td>
<td>60.22</td>
<td>16.48</td>
</tr>
<tr>
<td>Proposed method</td>
<td><strong>76.28</strong></td>
<td>66.92</td>
<td><strong>35.93</strong></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Method</th>
<th>ES-EN</th>
<th>IT-EN</th>
<th>TR-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>10</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Proposed method</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), ( \lambda = 1 )</td>
</tr>
<tr>
<td>Zhang et al. (2017), ( \lambda = 10 )</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
</tr>
<tr>
<td>Proposed method</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

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

Best accuracy (%)
Experiments

- Bilingual lexicon extraction
- 10 runs for each method
  \[ \Rightarrow \text{Best/average accuracy} \]
  \[ \Rightarrow \text{Successful runs (>5% accuracy)} \]
- (Easy) dataset by Zhang et al. (2017)
- (Hard) dataset by Dinu et al. (2016) + extensions

<table>
<thead>
<tr>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), ( \lambda = 1 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), ( \lambda = 10 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>45.40</td>
<td>47.27</td>
<td>1.62</td>
<td>36.20</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>45.27</td>
<td>0.07</td>
<td>0.07</td>
<td>35.47</td>
</tr>
<tr>
<td>Proposed method</td>
<td>48.53</td>
<td>48.47</td>
<td>33.50</td>
<td>37.60</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), ( \lambda = 1 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), ( \lambda = 10 )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>45.40</td>
<td>47.27</td>
<td>1.62</td>
<td>36.20</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>45.27</td>
<td>0.07</td>
<td>0.07</td>
<td>35.47</td>
</tr>
<tr>
<td>Proposed method</td>
<td>48.53</td>
<td>48.47</td>
<td>33.50</td>
<td>37.60</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>13.55</td>
<td>42.15</td>
<td>0.38</td>
<td>21.23</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>9.10</td>
<td>0.01</td>
<td>0.01</td>
<td>7.09</td>
</tr>
<tr>
<td>Proposed method</td>
<td>48.13</td>
<td>48.19</td>
<td>32.63</td>
<td>37.33</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Conneau et al. (2018), code</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Conneau et al. (2018), paper</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Proposed method</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Number of successful runs
Experiments

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mikolov et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Faruqui and Dyer (2014)</td>
</tr>
<tr>
<td></td>
<td>Shigeto et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Dinu et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Lazaridou et al. (2015)</td>
</tr>
<tr>
<td>5k dict.</td>
<td>Xing et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Smith et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2018)</td>
</tr>
</tbody>
</table>
## Experiments

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mikolov et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Faruqui and Dyer (2014)</td>
</tr>
<tr>
<td></td>
<td>Shigeto et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Dinu et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Lazaridou et al. (2015)</td>
</tr>
<tr>
<td>5k dict.</td>
<td>Xing et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Smith et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2018)</td>
</tr>
<tr>
<td>25 dict.</td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td>Init.</td>
<td>Smith et al. (2017), cognates</td>
</tr>
<tr>
<td>heurist.</td>
<td>Artetxe et al. (2017), num.</td>
</tr>
<tr>
<td>Supervision</td>
<td>Method</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>None</td>
<td>Conneau et al. (2018), code‡</td>
</tr>
<tr>
<td>Init. heurist.</td>
<td>Smith et al. (2017), cognates</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2017), num.</td>
</tr>
<tr>
<td>25 dict.</td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td>5k dict.</td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Xing et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Smith et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>proposed method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Conneau et al. (2018), paper‡</td>
</tr>
<tr>
<td></td>
<td>proposed method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Conneau et al. (2018), code‡</td>
</tr>
<tr>
<td></td>
<td>proposed method</td>
</tr>
</tbody>
</table>

Artetxe et al. (2016), Artetxe et al. (2017), Smith et al. (2017), Zhang et al. (2016), Zhang et al. (2017), λ = 1, λ = 10
## Experiments

<table>
<thead>
<tr>
<th>Supervision</th>
<th>Method</th>
<th>EN-IT</th>
<th>EN-DE</th>
<th>EN-FI</th>
<th>EN-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>5k dict.</td>
<td>Mikolov et al. (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Faruqui and Dyer (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shigeto et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dinu et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lazaridou et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xing et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smith et al. (2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artetxe et al. (2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 dict.</td>
<td>Artetxe et al. (2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Init. heurist.</td>
<td>Smith et al. (2017), cognates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Init. heurist.</td>
<td>Artetxe et al. (2017), num.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Zhang et al. (2017), $\lambda = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Zhang et al. (2017), $\lambda = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Conneau et al. (2018), code †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Conneau et al. (2018), paper †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Experiments

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

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

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

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

https://github.com/artetxem/vecmap